



Université des Antilles et de la Guyane

Faculté des Sciences Exactes et Naturelles

Année 2012
bibliothèque

N° attribué par la

/ / / / / / / / / / /

THESE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE des ANTILLES et de la GUYANE

Discipline : Biologie Santé

Présentée par

Véronique HILL

Phylogéographie mondiale des bacilles tuberculeux : contribution des outils moléculaires et bioinformatiques, et caractérisation des lignées génotypiques pour des études épidémiologiques et phylogénétiques

Directeur de thèse :
Dr. Nalin RASTOGI

Soutenue publiquement le 07 juin 2012

JURY PROVISoire

Professeur Christian PERRONNE	Rapporteur
Docteur Helmi MARDASSI	Rapporteur
Professeur Raymond CESAIRE	Examineur
Professeur Olivier GROS	Examineur
Docteur Antoine TALARMIN	Examineur
Docteur Nalin RASTOGI	Directeur de Thèse

THESE

Pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITE des ANTILLES et de la GUYANE

Discipline : Biologie Santé

Présentée par

Véronique HILL

**Phylogéographie mondiale des bacilles tuberculeux : contribution
des outils moléculaires et bioinformatiques, et caractérisation des
lignées génotypiques pour des études épidémiologiques et
phylogénétiques**

Directeur de thèse :
Dr. Nalin RASTOGI

Soutenue publiquement le 07 juin 2012

REMERCIEMENTS

Je tiens à remercier Dr. Nalin Rastogi, Directeur du laboratoire des mycobactéries et de la tuberculose, de m'avoir donné l'opportunité de faire cette thèse. Je salue l'impulsion qu'il a su donner à ce laboratoire de recherche.

Je remercie le précédent Directeur de l'Institut Pasteur de la Guadeloupe (IPG) Monsieur Ronald Perraut, l'actuel Directeur Monsieur Antoine Talarmin ainsi que le Directeur financier de l'Institut Monsieur Marcel Sigiscar, qui œuvrent pour que les différents travaux menés au sein de l'IPG se déroulent dans des conditions favorables à leur aboutissement.

Je remercie Dr. Thierry Zoio pour l'aide et l'assistance qu'il m'a apportées dans les différents projets menés dans le cadre de cette thèse. Je remercie également tous les autres membres de l'équipe ainsi que tous les stagiaires que j'ai côtoyés durant ces 3 dernières années.

Je remercie Pr. Christian Perronne, Dr. Helmi Mardassi, Pr. Raymond Césaire, Pr. Olivier Gros, Dr. Antoine Talarmin, de me faire l'honneur d'être membre de mon jury de thèse.

SOMMAIRE

INTRODUCTION

1. Introduction générale sur la tuberculose	13
1.1. Description de la tuberculose	13
1.1.1. De la phtisie à la tuberculose	13
1.1.2. Symptômes, transmission et facteurs de risque.....	15
1.1.3. Diagnostic bactériologique de la tuberculose : diagnostic indirect et diagnostic direct	16
1.1.4. Traitement de la tuberculose.....	18
1.1.4.1. Une multi-thérapie antibiotique pour la prise en charge des patients.....	18
1.1.4.2. Les antibiogrammes : étude de la sensibilité du bacille tuberculeux aux antibiotiques	21
1.1.5. La prévention de la tuberculose au moyen de la vaccination par le BCG (Bacille Calmette Guérin).....	23
1.2. Épidémiologie descriptive de la tuberculose	23
1.2.1. Incidence, prévalence et mortalité tuberculeuse dans le monde	24
1.2.2. L'Afrique et Asie du Sud-Est, zones à forte incidence selon l'OMS	25
1.3. Les préoccupations tuberculeuses actuelles	26
1.3.1. Une tuberculose pharmaco-résistante.....	26
1.3.2. Le Virus de l'Immunodéficience Humaine (VIH) et le Bacille de Koch (BK) : un tandem meurtrier	27
1.4. Les programmes de lutte antituberculeuse	27
1.4.1. La stratégie DOTS (Directly Observed Treatment Short-Course).....	27
1.4.2. La stratégie Halte à la tuberculose	28
2. Description des bacilles du genre <i>Mycobacterium</i>	29
2.1. Une paroi cellulaire caractéristique	29
2.1.1. Généralités sur la paroi cellulaire des bactéries à Gram +	29
2.1.2. Paroi cellulaire des mycobactéries.....	30
2.1.3. Paroi cellulaire et virulence mycobactérienne.....	31
2.2. Comportement des mycobactéries au sein de l'hôte sensible	32
2.2.1. Le granulome, un agrégat de macrophages pour l'isolation du site de l'infection	32
2.2.2. Tuberculose pulmonaire et extra pulmonaire	34
2.3. La classification des mycobactéries et description des diverses espèces mycobactériennes	35
2.3.1. Les principales méthodes de classification biologique	35
2.3.2. La notion d'espèce bactérienne.....	36
2.3.3. Diversité des espèces mycobactériennes et apport des marqueurs moléculaires	37
2.3.3.1. Généralités sur le chromosome mycobactérien.....	37
2.3.3.2. Le polymorphisme génétique : support de la différenciation des souches MTC au cours du temps	37
2.3.3.2.1. Les régions de différence chez <i>M. tuberculosis</i> , support des « Large Sequence Polymorphism » (LSP)	38
2.3.3.2.2. La séquence d'insertion IS6110.....	40
2.3.3.2.3. Le polymorphisme d'un seul nucléotide (SNP ou single-nucleotide polymorphism).....	41
2.3.3.2.4. Les loci CRISPR de <i>M. tuberculosis</i> support du spoligotypage	42
2.3.3.2.5. Le polymorphisme du nombre d'unités répétées VNTR (Variable Number Tandem Repeats)	44

2.3.4.	Description des espèces mycobactériennes	44
2.3.4.1.	Description des espèces appartenant au groupe des mycobactéries non tuberculeuses (MNT).....	45
2.3.4.2.	<i>Mycobacterium leprae</i> agent causal de la lèpre	46
2.3.4.3.	Description des espèces membres du complexe <i>Mycobacterium tuberculosis</i>	46
2.3.4.3.1.	Description de <i>Mycobacterium africanum</i>	47
2.3.4.3.2.	Description de <i>Mycobacterium caprae</i>	48
2.3.4.3.3.	Description de <i>Mycobacterium bovis</i>	49
2.3.4.3.4.	Description de <i>Mycobacterium microti</i>	50
2.3.4.3.5.	Description de <i>Mycobacterium pinnipedii</i>	51
2.3.4.3.6.	Description de <i>Mycobacterium mungi</i>	51
2.3.4.3.7.	Description de <i>Mycobacterium canettii</i>	52
2.3.4.3.8.	Description de <i>Mycobacterium tuberculosis</i>	52
3.	Notion d'évolution en biologie.....	53
3.1.	Les théories de l'évolution des êtres vivants.....	53
3.1.1.	Le transformisme de Lamark ou l'influence du milieu dans la complexification des êtres vivants.....	53
3.1.2.	La sélection naturelle de Charles Darwin et de Alfred Russel Wallace motrice de l'évolution biologique	54
3.1.3.	La biogéographie discipline introduite par Alfred Russel Wallace	55
3.2.	Les méthodes de reconstruction phylogénétique pour l'inférence des relations de parenté inter/intra-espèces	56
3.2.1.	La parcimonie de Wagner.....	57
3.2.2.	La méthode bayésienne : une approche probabiliste.....	57
3.2.3.	Les méthodes de distances	59
3.2.3.1.	Unweight Pair Group Method with Arithmetic mean (UPGMA).....	59
3.2.3.2.	Le Neighbor-Joining.....	60
3.2.3.3.	Le minimum spanning tree.....	60
4.	Evolution phylogénétique de MTC	61
4.1.	Histoire évolutive du complexe <i>Mycobacterium tuberculosis</i>	61
4.2.	Apport de la bioinformatique dans la reconstruction phylogénétique de MTC.	63
4.2.1.	La base de données de génotypes mycobactériens, outils indispensable pour comprendre l'évolution et la biogéographie de MTC	63
4.3.	Outils bioinformatiques pour l'acquisition de nouvelles connaissances sur le complexe <i>Mycobacterium tuberculosis</i>	65
4.3.1.	Les réseaux bayésiens, une technique de l'intelligence artificielle pour extraire les informations cachées dans les grandes bases de données	65
4.3.2.	Importance d'un modèle bayésien quant à la classification des souches du complexe <i>Mycobacterium tuberculosis</i>	66
5.	Les problématiques auxquelles doit répondre cette thèse	67

MATÉRIELS ET MÉTHODES

1.	Les méthodes de génotypage moléculaire utiles à la description des souches MTC.....	69
1.1.	Description de l'IS6110-RFLP (polymorphisme de longueur des fragments de restriction).....	69
1.1.1.	Principe général du polymorphisme de longueur des fragments de restriction	70
1.1.2.	Description du typage par IS6110-RFLP	70

1.2.	Description des techniques ayant recours à la PCR.....	71
1.2.1.	Le spoligotypage (Spacer Oligonucleotide Typing)	71
1.2.1.1.	Illustration du principe	71
1.2.1.2.	Description du protocole d'hybridation des produits PCR sur la membrane de nylon	73
1.2.1.3.	Traduction des résultats dans des formats adéquats pour leurs manipulations	73
1.2.2.	Technique pour la détection des insertions d'IS6110 dans le locus DR.....	74
1.2.2.1.	Intérêt	74
1.2.2.2.	Principe	74
1.2.3.	Les techniques basées sur l'amplification des minisatellites VNTR	76
1.2.3.1.	Description de la méthodologie	76
1.2.3.2.	Description des différents formats de loci considérés.....	77
1.2.4.	Mise en place d'une nouvelle technique pour la détection des délétions adjacentes au point d'insertion d'IS6110	78
1.2.4.1.	Principe	78
1.2.4.2.	Description du protocole	79
1.2.5.	Le génotypage basé sur le polymorphisme de longues séquences (LSP)	80
1.3.	PCR en temps réel et SNPs (Polymorphismes de nucléotide unique)	80
	Protocole PCR	81
2.	SITVITWEB : collection génotypique internationale de MTC.....	83
2.1.	Généralités sur la collection de données de SITVITWEB.....	83
2.2.	Description des diverses informations stockées dans SITVITWEB.....	83
2.3.	Comparaison avec les précédentes versions de la base ; SpoIDB1, SpoIDB2, SpoIDB3 et SpoIDB4	85
3.	Les logiciels de phylogénie utilisés.....	86
3.1.	MrBayes pour une phylogénétique selon la méthode bayésienne.....	86
3.2.	BioNumerics pour la construction phylogénétique selon la méthode du Minimum Spanning Tree (MST)	86
4.	Outils mathématiques utilisés.....	87
4.1.	Les réseaux bayésiens	87
4.1.1.	Définition.....	87
4.1.2.	Fonctionnement	87
4.1.3.	Choix de GeNIe pour l'apprentissage de structure et l'apprentissage des tables de probabilités	88
4.2.	La loi normale	88
4.2.1.	Généralités	88
4.2.2.	Renforcement de la prédiction du nombre de répétitions portées par chaque locus MIRU-VNTR grâce au Z-score	89
4.2.3.	Utilisation du Z-score cumulé pour la classification des souches MTC selon leur valeur de MIRU12.....	91
4.3.	L'index de discrimination de Hunter-Gaston (HGDI) pour évaluer la variabilité biologique	92

RÉSULTATS ET DISCUSSION

1.	Etat des lieux de la pertinence et des potentialités des données stockées dans la plus grande base de données génotypiques SITVITWEB (Article 1)....	93
1.1.	SITVITWEB : une mise à jour de la diversité génotypique des souches MTC circulant dans le monde.....	93

1.1.1.	Apport des différents marqueurs moléculaires dans la description des profils génotypiques	93
1.1.1.1.	Apport du spoligotypage	93
1.1.1.2.	Apport du typage basé sur les loci MIRU-VNTR (formats 12 loci et 5 loci)	95
1.2.	Intérêt du spoligotypage dans la description des profils génotypiques de MTC	97
1.2.1.	Mise à jour des lignées spoligotypiques de SpolDB4	97
1.2.1.1.	Méthodologie appliquée pour la mise à jour de la classification spoligotypique	97
1.2.1.2.	Présentation des prototypes	97
1.2.1.3.	Description de la procédure algorithmique de la classification bioinformatique	99
1.2.1.4.	Bilan de la classification des souches dans SITVITWEB	100
1.2.2.	Description de la phylogéographie de MTC à l'aide de la classification spoligotypique	100
1.2.2.1.	Phylogéographie mondiale de MTC	100
1.2.2.2.	Géospécificité des sous-lignées de MTC dans deux zones géographiques restreintes et distinctes	102
1.2.2.2.1.	Description du paysage génotypique de MTC au Mozambique (Article 2)	102
1.2.2.2.1.1.	Description de la collection d'isolats cliniques	102
1.2.2.2.1.2.	Description du paysage spoligotypique du Mozambique	103
1.2.2.2.1.3.	Mise en évidence de deux cas d'infection polyclonale parmi les 5 isolats cliniques Manu2 de l'étude	105
1.2.2.2.2.	Description du paysage génotypique de MTC de la République de Guinée-Bissau (Article 3)	106
1.2.2.2.2.1.	Description de la collection d'isolats cliniques	106
1.2.2.2.2.2.	Description du paysage spoligotypique de Guinée-Bissau	107
1.2.2.2.2.3.	Caractérisation de la lignée phylogénétique majeure en Guinée-Bissau	108
1.2.2.2.3.	Conclusion sur la géospécificité des lignées	109
1.3.	Les limites du spoligotypage	110
1.3.1.	Les profils spoligotypiques et les infections polyclonales (Article 4)	110
1.3.1.1.	Généralités sur les infections polyclonales	110
1.3.1.2.	Approche informatique pour la détection des profils spoligotypiques composites	110
1.3.1.3.	Description des profils spoligotypiques les plus concernés	113
1.3.1.3.1.	Les profils les plus exposés : SIT53, SIT100, SIT54, SIT42, SIT236	113
1.3.1.3.2.	Le groupe phylogénétique Manu : est-ce un artéfact ?	115
1.3.2.	Incapacités du spoligotypage à discriminer les souches de la lignée phylogénétique Beijing (Article 5)	117
1.3.2.1.	Description de la collection d'isolats cliniques de Sichuan en Chine	118
1.3.2.2.	Subdivision des clusters de spoligotype par les loci MIRU-VNTR	119
1.3.2.3.	Subdivision des isolats Beijing et non-Beijing par les résultats de l'IS6110-RFLP	119
1.3.2.4.	Conclusion sur la discrimination des souches Beijing	119
1.4.	Quel est le potentiel réel des loci VNTR dans l'épidémiologie et la phylogénie de la tuberculose ? (Article 1)	120
1.4.1.	La variabilité des loci VNTR à l'intérieur des clones spoligotypiques	120
1.4.2.	Hypothèse d'une variabilité spécifique des loci MIRU-VNTR au sein de différentes lignées phylogénétiques	121
2.	Utilisation de l'intelligence artificielle pour capturer les informations cachées dans la base de données SITVITWEB (Article 6)	123
2.1.	VNTR-TBbaves, une application WEB pour l'aide à la classification, à la complétude des données de typage manquantes et à l'attribution d'une origine géographique aux souches de MTC	124

2.1.1.	Intérêt des réseaux bayésiens pour MTC	124
2.1.2.	Choix des variables	124
2.1.3.	Description des échantillons utilisés pour construire et tester le modèle	125
2.2.	Evaluation de la pertinence de VNTR-TBbayes dans les prédictions.....	126
2.2.1.	Evaluation de la pertinence de VNTR-TBbayes dans l'assignation des lignées.....	126
2.2.1.1.	Seul le spoligotype est placé en évidence	126
2.2.1.2.	Seul le MIRU12 est placé en évidence	128
2.2.1.3.	Le spoligotype et le MIRU12 sont placés en évidence.....	128
2.2.1.4.	Conclusion sur la pertinence de VNTR-TBbayes dans l'assignation des lignées à partir des données issues du spoligotypage et/ou du typage des minisatellites MIRU-VNTR.....	128
2.2.2.	Evaluation de la pertinence de VNTR-TBbayes dans l'assignation des sous-lignées connaissant la lignée	129
2.2.3.	Evaluation dans la complétude des informations de typage	130
2.3.	Détermination par VNTR-TBbayes du profil MIRU-VNTR caractéristique de chaque lignée/sous-lignée phylogénétique	130
2.3.1.	Présentation des profils MIRU-VNTR ou nœuds phylogénétiques centraux.....	130
2.1.1.	Construction d'un arbre bayésien avec les différents nœuds centraux pour discuter des résultats de VNTR-TBbayes	132
2.2.	Evolution et perspectives de VNTR-TBbayes.....	135
2.2.1.	Apport de nouvelles variables au modèle bayésien.....	135
2.2.1.1.	Les variables potentielles.....	135
2.2.1.2.	Etude réalisée au Nigéria : le statut VIH, une variable dans un modèle bayésien pour MTC ? (Article 7).....	136
2.2.1.2.1.	Description de la collection d'échantillons de MTC de l'étude	137
2.2.1.2.2.	Description de la population de MTC au Nigéria chez les patient HIV+ ..	137
2.2.1.2.3.	Description de la population de MTC au Nigéria chez les patient HIV- ...	138
2.2.1.2.4.	Conclusion sur le statut VIH et les profils génotypiques	138
3.	Une classification basée sur les données de typage MIRU-VNTR pour une phylogénie plus robuste de MTC (Article 8).....	141
3.1.	Définition d'un prototype MIRU-VNTR à l'aide d'une étude réalisée en Inde (Article 9).....	141
3.1.1.	Épidémiologie moléculaire des isolats MTC à Kerala en Inde à l'aide des techniques <i>IS6110-RFLP</i> , spoligotypage ⁴³ et du typage par MIRU-VNTR.....	141
3.1.1.1.	Description de la collection de souches de Kerala en Inde.....	142
3.1.1.2.	Description du paysage génotypique à Kerala en Inde	142
3.1.1.2.1.	Description des spoligotypes répertoriés	142
3.1.1.2.2.	Analyse des résultats de typage MIRU-VNTR	145
3.1.2.	Illustration de l'importance des profils MIRU-VNTR centraux prédits par VNTR-TBbayes	146
3.1.3.	Méthodologie utilisée pour l'identification des lignées selon les minisatellites MIRU-VNTR et la classification des patterns MIRU12	149
3.2.	Apport de nouvelles données de typage pour compléter la base SITVIT.....	149
3.2.1.	Profils génotypiques caractéristiques de la côte Ouest africaine dans une étude menée au Nigéria (Article 10)	150
3.2.1.1.	Description de la collection de souches.....	150
3.2.1.2.	Description des profils génotypiques au Nigéria	150
3.2.2.	Investigation des loci MIRU-VNTR sur la côte Est africaine dans une étude menée au Mozambique (Article 8)	152
3.2.3.	Présentation et analyse des nouvelles lignées basées sur 12 loci MIRU-VNTR.....	153
3.2.3.1.	Présentation des 41 signatures génotypiques identifiées.....	153
3.2.3.2.	Confrontation de cette nouvelle classification aux autres schémas de classification publiés	155
3.2.3.2.1.	Reclassification des profils de la base de données MIRU-VNTRplus	156
3.2.3.2.2.	Reclassification des profils de la base de données SITVIT	157

3.2.3.3.	Superposition de cette nouvelle classification à la classification de Gagneux basée sur les LSPs	159
3.2.4.	Hypothèse évolutive et relations phylogénétiques entre les lignées nouvelles définies chez MTC	159
3.2.4.1.	Construction d'une phylogénie bayésienne	159
3.2.4.2.	Identification de deux groupes phylogénétiques très importants chez les souches modernes PGG2/3	161
3.2.4.3.	Identification du groupe phylogénétique présentant la délétion RD(Rio)	163
3.2.4.3.1.	Reclassification de profils MIRU-VNTR d'étude déjà publiée	164
3.2.4.3.2.	Description du groupe phylogénétique RD(Rio)- dans le schéma de classification basé sur les 12 loci MIRU-VNTR	164
3.2.5.	Discordances entre la classification spoligotypique et celle des MIRU-VNTR	165
3.2.5.1.	Exemples de discordances	165
3.2.5.2.	Implication probable de l'IS6110 dans le polymorphisme du locus DR	166
3.2.5.2.1.	Mise en évidence d'une forte présence de d'IS6110 dans le locus DR ...	166
3.2.5.2.2.	Mise en évidence de délétions adjacentes aux zones d'insertion de l'élément transposable IS6110	168
3.2.5.3.	Comment expliquer les discordances entre la classification spoligotypique et celle basée sur les MIRU-VNTR	170
4.	Conclusion générale et perspectives	171
4.1.	Intérêt global d'une méthodologie de classification basée sur les minisatellites VNTR	171
4.1.1.	Intégration des lignées génotypiques de la classification MLVA au modèle VNTR-TBbayes	171
4.1.2.	Réévaluation de la classification spoligotypique sur la base de la classification MIRU-VNTR	172
4.1.3.	Un exemple d'évolution convergente dans le locus DR de <i>Mycobacterium tuberculosis</i> (Fenner et al., 2011)	177
4.2.	Le passage à une classification basée sur les 24 loci MIRU-VNTR : est-ce un impératif ?	178
4.3.	Illustration de la géospécificité des lignées basées sur les MIRU-VNTR, relation avec l'histoire du peuplement du monde (Out of Africa)	179
4.4.	La capacité des loci VNTR à décrire les lignées phylogénétiques de MTC, sous-entendrait-elle un rôle de ces minisatellites chez les mycobactéries ?	184
4.4.1.	Analyse bioinformatique des séquences de 12 loci MIRU-VNTR	184
4.4.2.	Les loci MIRU-VNTRs seraient-ils les supports d'un mécanisme de régulation traductionnelle par le recodage du codon stop et le décalage de cadre de lecture chez les souches de MTC ?	186
4.4.2.1.	Quel serait l'intérêt des codons stop portés par les minisatellites MIRU-VNTR ?	186
4.4.2.2.	Mécanisme d'insertion de la sélénocystéine (Sec) au cours de la traduction	187
4.4.2.3.	Structure et fonction de l'acide aminé sélénocystéine (Sec)	188
4.4.2.4.	Compétition entre l'insertion de Sec et la terminaison	188
4.4.2.5.	Quel serait l'intérêt des séquences Shine-Dalgarno portées par les VNTRs ? (réponse apportée par l'article Namy et al., 2004)	189
4.4.2.6.	Conclusion sur une éventuelle fonction des minisatellites MIRU-VNTR chez les souches MTC	191

RÉFÉRENCES

ANNEXES

PUBLICATIONS SCIENTIFIQUES ISSUES DE CE TRAVAIL CITÉES DANS L'ORDRE D'APPARITION

Article 1

Demay C, Liens B, Burguière T, **Hill V**, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N (2011) SITVITWEB – A publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infect Genet Evol.

Article 2

Viegas SO, Machado A, Groenheit R, Ghebremichael S, Pennhag A, Gudo PS, Cuna Z, Miotto P, **Hill V**, Marrufo T, Cirillo DM, Rastogi N, Källenius G, Koivula T. Molecular diversity of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in Mozambique. BMC Microbiol. 2010 Jul 21;10:195.

Article 3

Groenheit R, Ghebremichael S, Svensson J, Rabna P, Colombatti R, Riccardi F, Couvin D, **Hill V**, Rastogi N, Koivula T, Källenius G. The Guinea-Bissau family of *Mycobacterium tuberculosis* complex revisited. PLoS One. 2011 Apr 20;6(4):e18601.

Article 4

Lazzarini LC, Rosenfeld J, Huard RC, **Hill V**, Lapa E Silva JR, Desalle R, Rastogi N, Ho JL. *Mycobacterium tuberculosis* spoligotypes that may derive from mixed strain infections are revealed by a novel computational approach. Infect Genet Evol. 2011 Sep 5.

Article 5 (soumis)

Yuding Zhao, Qin Feng, Chao Zheng, Yekun Sun, Qiang Luo, Tao Luo, **Véronique Hill**, Nalin Rastogi, and Qun Sun. Use of 24-locus MIRU-VNTR and IS6110 Restriction Fragment Length Polymorphism for Differentiating Spoligotype-defined Clusters of Drug-resistant *Mycobacterium tuberculosis* Clinical Isolates from Sichuan in China. Journal of Clinical Microbiology.

Article 6 (soumis)

Véronique Hill, Thierry Zozio, David couvin, Nalin Rastogi. VNTR-TBbayes: Using artificial intelligence to capture hidden knowledge in huge international *Mycobacterium tuberculosis* genotyping databases. PLoS ONE.

Article 7

Cadmus S, **Hill V**, van Soolingen D, Rastogi N. Spoligotype profile of *Mycobacterium tuberculosis* complex strains from HIV-positive and –negative patients in Nigeria: a comparative analysis. J Clin Microbiol. 2011 Jan;49(1):220-6. Epub 2010 Nov 3.

Article 8 (accepté)

Véronique Hill, Thierry Zozio, Syndia Sadikalay, Sofia Viegas, Elisabeth Streit, Gunilla Kallenius, Nalin Rastogi. MLVA based classification of *Mycobacterium tuberculosis* complex lineages for a robust phylogeographic snapshot of worldwide molecular diversity. PLoS ONE.

Article 9 (soumis)

Biljo V Joseph, Smitha Soman, Indulakshmi Radhakrishnan, **Véronique Hill**, R. Ajay Kumar, Nalin Rastogi, Sathish Mundayoor. Molecular epidemiology of *Mycobacterium tuberculosis* isolates using IS6110-RFLP, Spoligotyping and MIRU-VNTRs from tuberculosis patients in Kerala, India. PLoS ONE.

Article 10

Thumamo BP, Asuquo AE, Abia-Bassey LN, Lawson L, **Hill V**, Zozio T, Emenyonu N, Eko FO, Rastogi N. Molecular epidemiology and genetic diversity of *Mycobacterium tuberculosis* complex in the Cross River State, Nigeria. Infect Genet Evol. 2011 Aug 22.

RÉSUMÉ

La tuberculose, maladie très ancienne, est plus que jamais présente sur la scène sanitaire mondiale avec 9 millions de nouveaux cas par an. Cette maladie, qui demeure une cause majeure de mortalité, est l'un des problèmes les plus difficiles à résoudre en santé publique et à l'échelle internationale. Les divers programmes de lutte, lancés par l'Organisation Mondiale de la Santé (OMS), pour prévenir et garantir la guérison des patients tuberculeux, se heurtent à des difficultés qui hypothèquent leur efficacité. En effet, l'expansion du virus de l'immunodéficience humaine (VIH) et son influence néfaste sur la susceptibilité à la tuberculose, l'accroissement rapide de la tuberculose multi-résistante, la pauvreté ainsi que la croissance démographique, sont un ensemble de facteurs préjudiciables à l'application des mesures préventives et curatives. Les contextes économiques difficiles conduisant indubitablement au cumul de ces facteurs, ont fait une césure entre les pays riches et les pays pauvres quant à l'incidence et à l'évolution de la tuberculose à l'échelle mondiale.

Pour atténuer la disparité des résultats des luttes antituberculeuses, les méthodologies employées en épidémiologie de la tuberculose doivent s'inscrire dans une démarche d'efficacité. Les marqueurs moléculaires, indispensables à l'identification et à la classification des souches appartenant au complexe *Mycobacterium tuberculosis* (MTC) ont chacun des potentialités qui leur sont propres dont il convient de cerner pour une utilisation adéquate dans les études épidémiologiques et phylogénétiques. Le but de cette thèse consiste à extraire des connaissances nouvelles des résultats de génotypage stockées dans les bases de données, connaissances relatives au polymorphisme du support génomique pour estimer la contribution de chaque marqueur dans la classification de MTC. Le recours à différentes méthodes de reconstruction phylogénétique et de modélisation des données fut indispensable. À la lumière des résultats obtenus, cette thèse propose une nouvelle méthodologie de classification, pour l'établissement d'une phylogéographie robuste de MTC en relation avec l'histoire des migrations humaines qui aurait débuté en Afrique subsaharienne

Chaque méthode de génotypage ayant une efficacité et un coût propres, qu'il convient de mesurer par rapport aux moyens disponibles et aux résultats escomptés, il est donc intéressant de renforcer la connaissance des marqueurs et de proposer différentes alternatives menant à la classification des mycobactéries. Ainsi, les résultats obtenus dans cette thèse apportent plus de cohérence aux solutions proposées pour la caractérisation des clones de MTC circulant dans le monde.

INTRODUCTION

1. Introduction générale sur la tuberculose

1.1. Description de la tuberculose

1.1.1. De la phtisie à la tuberculose

La tuberculose est une maladie infectieuse qui touchait déjà les populations humaines de l'antiquité. En effet, des lésions caractéristiques de l'infection furent identifiées sur les fragments de colonne vertébrale de momies égyptiennes datant de 2400 ans av. J.-C. Dans ces écrits, Hippocrate le Grand (460-370 av. J.-C), précurseur de la médecine, nomme la maladie « phtisie » du fait de sa symptomatologie conduisant les malades à un état de faiblesse quasiment toujours fatal où ils semblaient se consumer, (phtisie provenant du mot grec consommation). La phtisie est alors décrite comme une infection broncho-pulmonaire et pleurale qui se caractérise par une toux persistante, une langueur, un amaigrissement et la présence de sang dans les crachats. Si l'investigation des symptômes n'est pas sujette à la controverse, il n'en est pas de même pour les causes de la maladie. Certains attribuent à la phtisie un caractère héréditaire alors que d'autres comme Aristote (IV^{ème} siècle av. J.-C) optent pour une nature contagieuse. Le médecin italien Girolamo Fracastoro prit parti pour l'origine contagieuse de l'infection dans son oeuvre *De Contagione et Contagiosis Morbis* (De la contagion et des maladies infectieuses, 1546). Selon lui, les maladies infectieuses auraient pour vecteurs des organismes invisibles à l'œil nu, qu'il nomme *seminaria contagionis*. Il soutient que ces microorganismes vivants seraient transmissibles d'un individu à un autre par contact direct, par l'air ou des objets souillés ou encore à cause d'un caractère héréditaire. La méconnaissance, relative à l'étiologie et à la physiopathologie de la maladie, a longtemps persisté, mais les connaissances biologiques croissantes couplées aux avancées en microscopie optique (avec Antoni van Leeuwenhoek, 1632-1723) ont lentement levé le voile sur cette maladie. Ce n'est qu'à partir du XVII^{ème} siècle que le médecin et anatomiste néerlandais, Franciscus de le Boë Sylvius, apporta la première description anatomique en attribuant la présence de nodules pulmonaires ou « tubercules » à la maladie. En 1720, Benjamin Marten, dans son ouvrage *A New Theory of Consumptions: More Especially of a Phthisis or*

Consumption of the Lungs souligne le caractère contagieux de la phtisie en mettant en cause des « animacules », nom antérieurement donné à l'ensemble des microorganismes observés en microscopie optique par Antoni van Leeuwenhoek. En 1733, Pierre Desault (chirurgien français) affirme que les tubercules sont la caractéristique principale de la maladie. Il étend donc la description de la phtisie à la forme pulmonaire déjà admise, il ajoute l'atteinte extra-pulmonaire puisque selon lui, «les causes de la phtisie et des écrouelles ont une parfaite ressemblance, soit dans leur naissance, soit dans leurs progrès de manière que l'on peut appeler la phtisie l'écrouelle du poumon». En 1810, Gaspard Laurent Bayle identifia six formes de phtisie (enkystée ou granulaire, miliaire, calcifiée, ulcéreuse, mélanique et cancéreuse) à l'issue d'observations faites sur neuf cents cas d'autopsie. En 1819, Laennec distingue la phtisie des autres maladies pulmonaires notamment celles dues à des tumeurs malignes pulmonaires. S'il reconnaît l'origine infectieuse de la phtisie, il n'admet pas pour autant sa contagiosité. C'est enfin en 1839 que le médecin allemand Johann Lukas Schönlein attribua le nom tuberculose à la maladie et unifia dans ce terme toutes ses manifestations cliniques. En 1843 (ou 1846) Philipp Klencke a été le premier à démontrer expérimentalement le caractère transmissible de la tuberculose. Un an plus tard (1844), Jakob Henle soutient également la contagiosité mais en ajoutant que la maladie n'était vérifiable que dans certaines circonstances. L'omniprésence de la maladie au XIX^{ème} siècle (une personne sur quatre en serait atteinte en Europe), est telle que de nombreuses œuvres littéraires en font allusion mais dans des situations où les sentiments atteignent leur paroxysme. En effet, dans la croyance populaire, la tuberculose touchait des êtres fragiles que la passion consumait : l'héroïne de la Dame aux Camélias, célèbre roman d'Alexandre Dumas fils écrit en 1848, est un bel exemple de cette vision romantique. Dans ce récit d'Alexandre Dumas, joliment interprété par l'actrice Isabelle Huppert, il était coutume de boire du lait de bovin en pensant pouvoir guérir de la tuberculose. À l'époque aucun traitement ne s'était imposé à part les cures d'air que le médecin allemand Hermann Brehmer fut le premier à préconiser (1854). De cette idée, se développèrent les centres de cure spécialisés dans le traitement de la tuberculose, les sanatoriums, qui prônaient le repos et la salubrité de l'air. Dans les années 60, le médecin Jean-Antoine Villemin parvint à reproduire la tuberculose chez des animaux en leur inoculant des tissus humains infectés. Il en déduit qu'un microbe était en cause, et par d'autres expériences, il

démontra que la transmission se fait par voie aérienne. C'est finalement Robert Koch qui apporta la preuve irréfutable qu'un microorganisme est le vecteur de la tuberculose. Lors d'une conférence spectaculaire qui s'est tenue à Berlin le 24 mars 1881, il dévoila la bactérie qui s'était soustraite de la connaissance de tant de chercheurs. A l'aide de techniques de coloration spéciales, il mit en évidence dans des tissus tuberculeux, des bâtonnets qu'il appela bacilles tuberculeux ou *Mycobacterium tuberculosis*.

1.1.2. Symptômes, transmission et facteurs de risque

Souvent, le premier contact entre une personne et le bacille tuberculeux (couramment appelé bacille de Koch ou BK) ne déclenche pas symptôme. Il s'agit ici de la primo-infection. On observe cependant la formation d'un chancre tuberculeux (ou chancre de primo-infection) à l'issue d'une radiographie pulmonaire. Ce chancre correspond à une masse remplie de BK qui prend l'aspect de lait caillé pour ensuite se calcifier. Le caséum ainsi formé est une nécrose quasi-pathognomonique de la tuberculose.

Lorsque les symptômes se manifestent, on observe souvent une toux (sèche la plupart du temps) et un état de fatigue. Deux à trois semaines après la primo-infection, la toux s'accompagne de glaires qui peuvent être mêlées de sang. La liste des symptômes est alors plus longue : douleurs thoraciques, détresse respiratoire, essoufflement lors de l'effort, perte d'appétit et de poids, fièvre, sueurs nocturne, fatigue et altération de l'état général.

La transmission la plus courante se fait par voie aérienne. Une personne en contamine une autre par des gouttelettes de salives chargées de BK. Seules les personnes infectées présentant des bactéries dans leurs expectorations sont contagieuses. Cela implique donc seulement les personnes présentant une tuberculose pulmonaire. Les formes extra-pulmonaires (osseuse, ganglionnaire, urogénitale, pleurale, méningée) sont rarement contagieuses. La contagiosité ne concerne pas les personnes présentant une tuberculose latente (tuberculose asymptomatique qui s'oppose à la tuberculose maladie) qui concerne neuf personnes infectées sur dix.

Le mode de transmission de la tuberculose fait de la proximité avec un individu tuberculeux, le pilier de tous les facteurs de risque. Ainsi, tout lieu ou situation qui

nécessite une promiscuité entre individus peut être un foyer de transmission : la famille, les collectivités, les maisons carcérales. De plus, certaines personnes sont plus vulnérables que d'autres. C'est le cas pour les enfants, les malades immunodéprimés, les personnes suivant un traitement immunosuppresseurs ou infectées par le virus de l'immunodéficience humaine (VIH). Le diagnostic de la maladie est donc une étape essentielle pour l'adoption de comportements appropriés pour limiter la transmission, aussi bien pour une prise en charge thérapeutique adaptée afin d'éviter l'expansion de l'infection à d'autres organes.

1.1.3. Diagnostic bactériologique de la tuberculose : diagnostic indirect et diagnostic direct

Un examen bactériologique doit être réalisé dans le cas d'une suspicion de tuberculose à la suite d'un examen clinique. Le dépistage de la tuberculose commence par une intradermo-réaction (IDR ou test de Mantoux) à la tuberculine. Ce test tuberculinique cutané qui constitue le diagnostic indirect révèle simplement un contact antérieur avec un bacille tuberculeux. L'injection de tuberculine (filtrat purifié de BK cultivés en milieu synthétique) provoque une réaction immunitaire induite par les antigènes mycobactériens visible grâce à une induration.

Lors de la première visite médicale d'un individu à risque (ayant eu un en contact antérieur avec un tuberculeux), le test tuberculinique est réalisé (figure 1). La suite de la procédure dépend de la taille de l'induration. Si celle-ci est supérieure ou égale à 10 mm, le test sera complété par une radiographie pulmonaire. Une radiographie est aussi indispensable pour les sujets en bas âge (inférieur à cinq ans), ou encore pour les immunodéprimés, puisque pour eux un test négatif n'écarte pas la possibilité d'une tuberculose évolutive. De plus, une chimioprophylaxie est prescrite sur les deux mois d'intervalle avec la deuxième visite médicale chez les sujets de moins de 15 ans ou toute autre personne en contact avec une personne récemment infectée. Ce délai est nécessaire pour détecter une tuberculose si le premier test était précoce. Ainsi, toute personne présentant un test tuberculinique négatif lors de la première visite est invitée à refaire ce dernier après deux mois. Le traitement chimioprophylactique pourra être arrêté si le deuxième test est négatif chez les moins de 15 ans. Les patients présentant un test positif à l'issue de la seconde visite seront traités si le virage du test IDR est récent (moins de deux ans). Un test positif

accompagné d'une radiographie normale et d'une absence de signes cliniques caractéristique d'une tuberculose latente. Dans ce cas, l'administration d'une chimioprophylaxie dépendra de la situation épidémiologique du patient. Elle sera dans tous les cas proposée aux enfants et aux malades immunodéprimés.

Il est important de préciser que la vaccination par le BCG (Bacille Calmette Guérin) peut donner un résultat positif à l'IDR. La positivité du test sera donc interprétée en fonction de l'antériorité du patient.

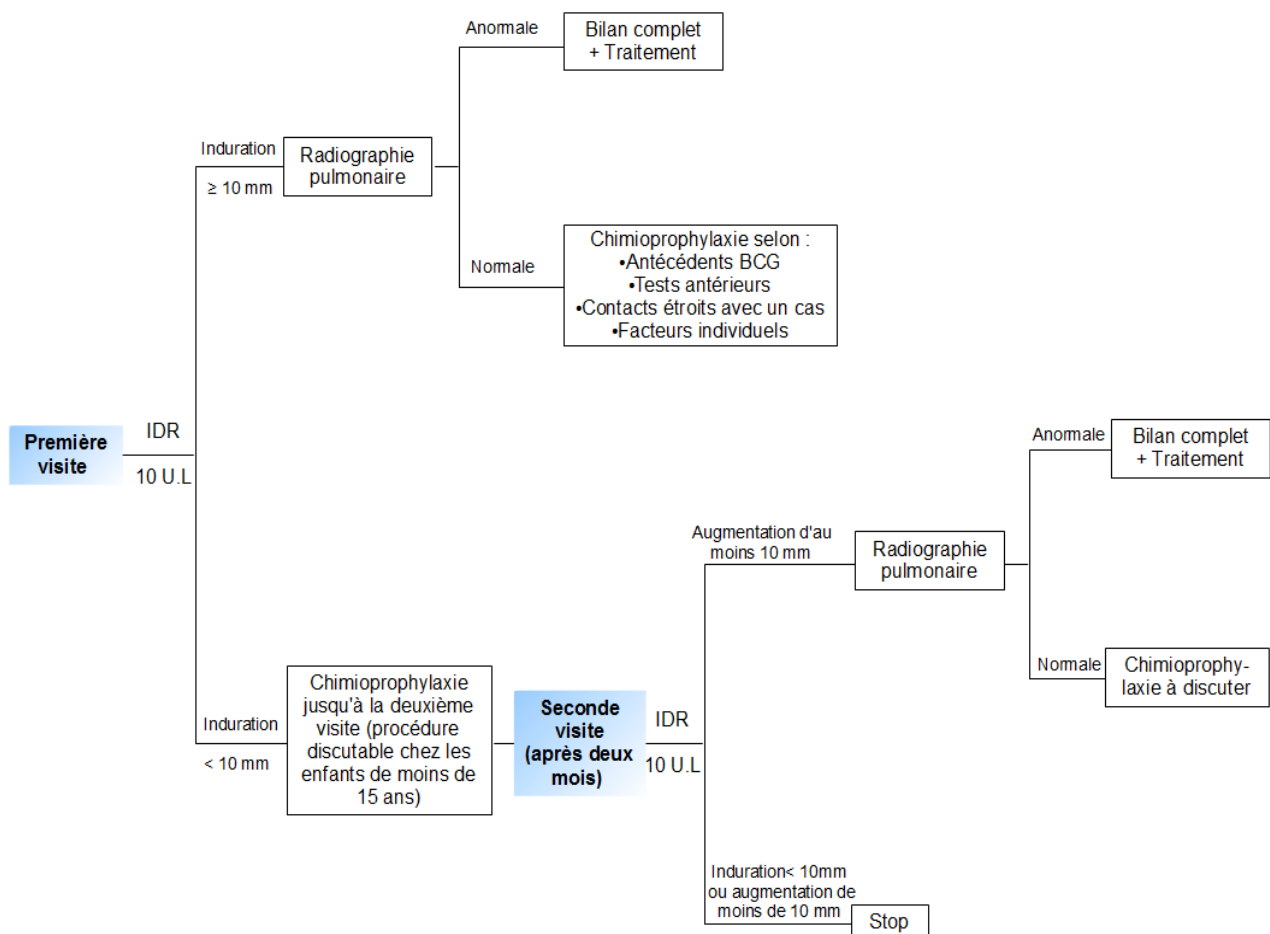


Figure 1. Schéma relationnel de la prise en charge du sujet depuis la première visite médicale jusqu'au début du traitement antibiotique.

Le diagnostic direct, qui repose sur l'examen cyto bactériologique des prélèvements pathologiques (crachats, liquide gastrique, urine etc.), est réalisé sur 3 jours consécutifs et accompagné d'un examen direct et de la mise en culture de l'éventuel bacille tuberculeux. L'utilisation du liquide gastrique (obtenu par tubage

gastrique) est une alternative pour les patients présentant une expectoration difficile. Le prélèvement de crachat ou de liquide gastrique est requis pour les cas de tuberculose pulmonaire. Cependant, la suspicion d'une tuberculose extra-pulmonaire requiert des prélèvements aux zones sièges de l'infection.

L'examen direct est un frottis coloré par la méthode de *Ziehl-Nielsen* qui révèle les bacilles acido-alcool-résistants (BAAR) par une coloration rose. Un frottis positif invoque la présomption accrue de tuberculose. Positive ou négative, cet examen microscopique est suivi par la mise en culture en milieu solide (Löwenstein-Jensen ou Middlebrook 7H11) ou liquide (Middlebrook 7H9), étape nécessaire pour l'établissement du diagnostic de certitude. Tout prélèvement pathologique doit être décontaminé avant la mise en culture, vu le caractère peu sélectif de ce dernier. La décontamination se fait généralement avec les méthodes à l'acétyl-cystéine ou à lauryl sulfate de sodium, antiseptiques auxquels les mycobactéries responsables de la tuberculose sont peu sensibles contrairement aux autres bactéries. Le diagnostic définitif peut donc être prononcé après 5 à 10 jours en milieu liquide, et environ 1 mois en milieu solide.

1.1.4. Traitement de la tuberculose

1.1.4.1. Une multi-thérapie antibiotique pour la prise en charge des patients

Tout patient pour lequel le diagnostic de la tuberculose est établi, doit bénéficier d'un traitement approprié. Si celui-ci n'a jamais été traité antérieurement, il convient de lui prescrire des antituberculeux de première ligne. Pour éviter le développement de souches résistantes aux antibiotiques il est judicieux de ne pas utiliser une monothérapie (terme relatif à l'usage d'un seul antibiotique). En effet, la multi-thérapie est préférée non seulement pour l'atténuation du risque d'émergence de résistance, mais également pour l'élargissement du spectre d'activité et le renforcement de l'effet bactéricide. La première phase du traitement comporte donc l'administration de 4 antituberculeux ; éthambutol (ETB), isoniazide (INH), pyrazinamide (PZA) et rifampicine (RIF) pendant deux mois. Au cours de la seconde étape, l'isoniazide et la rifampicine sont administrés pendant quatre mois.

Antibiotique	Structure moléculaire	Protéines cibles	Gène de la protéine cible	Mode d'action de l'antibiotique
Rifampicine (RIF) (Telenti et al., 1993), (Zhang et al., 1992)		Sous-unité β de l'ARN polymérase	rpoB 95% des souches RIF ^{res.}	Blocage de la transcription par fixation sur la protéine cible
Isoniazide (INH) (Zhang et al., 1992), (Rouse et al., 1995a), (Zhang, 1993), (Jacobs, 1994)		Catalase-péroxydase (haut niveau de résistance)	katG 50% à 70% des souches INH ^{res.}	Inhibition de : - l'activité catalase-péroxydase → augmentation des radicaux libres - la biosynthèse des acides mycoliques
(Banerjee et al., 1994), (Rouse et al., 1995b), (Veziaris et al., 2005)		Énoyl-ACP réductase (bas niveau de résistance)	inhA 20% à 30% des souches I NH ^{res.}	Fixation à la réductase et inhibition de la synthèse d'acides mycoliques à longues chaînes (Fas2) → Faiblesse de la paroi bactérienne
(Ristow et al., 1995), (Sherman et al., 1996), (Veziaris et al., 2005), (Zhang et al., 1992)		Alkyl hydro-péroxyde réductase C	ahpC 15% des souches INH ^{res.}	Inhibition des défenses contre le stress oxydatif
Pyrazinamide (PZA) (Sreevatsan et al., 1997a)		Pyrazinamidase/ Nicotinamidase	pncA 70% à 90% des souches PZA ^{res.}	Inhibition de la synthèse des acides gras à courte chaîne (Fas1), les mycolates.
Ethambutol (EMB) (Telenti et al., 1997), (Sreevatsan et al., 1997b)		Arabinosyl transférase	embB 65% des souches EMB ^{res.}	Inhibition de la formation liaison arabinose-galactane (mycolates).
Streptomycine (STR) (Honore et al., 1993), (Ramaswamy et al., 1998)		ARNr sous-unité 16S	rrs	Inhibition de l'ARNm
(Finken et al., 1993)		Protéine ribosomale S12	rpsL 65% à 75% des souches STR ^{res.}	Introduction d'erreurs de lecture
Fluoroquinolone (FQ) (Takiff et al., 1994)		ADN gyrase: sous unité A	gyrA	Fixation de la FQ et inhibition de l'ADN gyrase

Tableau 1. Les antibiotiques antituberculeux, leurs cibles et mécanismes de résistance. Tableau tiré du livre Zozio T, « *Diversité des bacilles tuberculeux circulant sur la planète* ». (Allemagne : Editions Universitaires Européennes, 2011).

Les antituberculeux induisent généralement des effets secondaires plus ou moins désagréables (coloration des urines, insuffisance rénale aiguë, hépatite médicamenteuse, troubles neuro-psychiques).

Cependant, les quatre antituberculeux ETB, INH, PZA et RIF sont généralement bien tolérés et leurs effets secondaires restent bénins. Le tableau 1 montre une description des principaux antibiotiques, leurs cibles et leurs mécanismes de résistance.

Quelques informations descriptives sur des antibiotiques de premières lignes :

L'**isoniazide** ou isonicotinyldiazine (INH), est un composé organique faisant partie des antituberculeux de première ligne. Découvert en 1912, il fut administré oralement pour le traitement de la tuberculose dès 1951. Son action antimycobactérienne s'explique par l'inhibition de la synthèse de l'acide mycolique. L'activation de l'isoniazide qui est un promédicament, nécessite l'intervention de la catalase peroxydase présente chez *M. tuberculosis* et codée par le gène *katG*. Cette catalase-péroxydase oxyde l'INH, ce qui aboutit à la synthèse d'un métabolite actif provoquant la mort cellulaire. Un second mécanisme serait proposé, il concerne l'inhibition de la synthèse de la protéine *InhA* intervenant dans la synthèse des acides mycoliques. Les résistances s'expliquent par deux mécanismes distincts. Le premier concerne les mutations sur le gène *katG* condamnant l'INH à rester inactif. On retrouve cette mutation chez 50% des souches résistantes à l'INH. Le niveau de résistance est défini par une concentration minimale inhibitrice (CMI) supérieure à 1 mg/L. Le second mécanisme responsable de la résistance à l'INH concerne les mutations du gène *inhA*, observées chez 10 à 30 % des souches résistantes. Dans ce dernier cas la CMI est inférieure à 1 mg/L.

La **rifampicine** appartient à la famille des rifamycines. La rifampicine bloque l'ARN polymérase. Les résistances primaires de *M. tuberculosis* sont dues à des mutations du gène *rpoB*, gène codant la sous-unité bêta de l'ARN polymérase. Cette mutation est observable chez la majorité des souches résistantes. Celles-ci présentent une CMI élevée, soit supérieure à 64 mg/L.

L'**éthambutol** a un effet bactériostatique sur les mycobactéries. Il entrave la formation de la paroi cellulaire en inhibant l'arabinosyl transférase. Cette enzyme est nécessaire à la synthèse des arabinogalactanes faisant partie du complexe mycolyl-arabinogalactane-peptidoglycane localisé dans la paroi cellulaire. Cela entraîne une augmentation de la perméabilité de la paroi cellulaire. Les arabinosyl transférases

sont codées par l'opéron *emb*. Les bactéries résistantes à l'ETB présentent des mutations sur les gènes *embB* (Garg et al., 2002).

Le **pyrazinamide** est apparentée au nicotinamide et est actif à pH acide. La pyrazinamidase, codée par le gène *pncA*, transforme le pyrazinamide en acide pyrazinoïque qui inhibe l'activité de l'acide gras synthase ou FAS (Zimhony et al., 2007). Il en résulte une baisse de la synthèse des acides gras et de façon inéluctable un blocage de la croissance bactérienne. L'acide pyrazinoïque se fixe à la protéine ribosomale S1 (gène *rs1/rpsA*) de la - 30S et empêche la trans-translation. Les mutations portées par les gènes *pncA* et *rpsA* sont donc responsables de la résistance à la pyrazinamide. Cette antituberculeux est préconisé dans le traitement de la tuberculose latente.

Le contrôle au préalable de la sensibilité aux antibiotiques de première ligne est indispensable à tout patient pour lequel on suspecte une infection par une souche résistante ou qui a bénéficié d'un précédent traitement, ou présentant une tuberculose chronique ou encore réfractaire au traitement (culture positive après trois mois de traitement). Ce test est réalisé par un antibiogramme. Les souches résistantes à la fois à l'INH et à la RIF sont qualifiées de souches multirésistantes ou MDR (Multi-Drug Resistant). En présence d'une souche MDR, on évalue la sensibilité aux antibiotiques de deuxième ligne (amikacine, kanamycine, capréomycine, ofloxacine, ciprofloxacine, éthionamide, prothionamide, rifabutine, cyclosérine, acide para-aminosalicylique). La tuberculose ultra-résistante ou eXtensively Drug-Resistant (XDR) met en cause une souche tuberculeuse MDR et à au moins un antibiotique de la famille des fluoroquinolones (ciprofloxacine, ofloxacine) et au moins un des trois médicaments injectables de deuxième ligne (capréomycine, kanamycine ou amikacine). Le traitement de la tuberculose MDR et XDR repose sur l'administration d'antibiotiques de deuxième intention choisis selon les résultats des antibiogrammes.

1.1.4.2. Les antibiogrammes : étude de la sensibilité du bacille tuberculeux aux antibiotiques

Les antibiogrammes se font sur milieu solide ou liquide de façon directe (à partir du culot de centrifugation d'un prélèvement biologique) ou indirecte (à partir d'une souche isolée sur un milieu solide). Une première lecture sera faite après une

vingtaine de jours et une seconde au quarantième jour afin de calculer le pourcentage de bacilles résistants. Ce calcul est basé sur la méthode des proportions. Celle-ci est évaluée en comptant le nombre de colonies présentes dans le milieu témoin et celles présentes dans les milieux avec antibiotique (bacilles résistants). La proportion de souches résistantes répond à la formule suivante : *Nombre de colonies résistantes / Nombre de colonies sur milieu témoin x 100*, Lorsqu'elle est inférieure à une proportion dite « proportion critique » on est en présence d'une souche sensible et au-delà, la souche est résistante. Par exemple cette proportion est de 1% pour l'INH, l'ETB, la RIF et de 10% pour le PZA.

Les antibiogrammes peuvent aussi être réalisés sur un milieu solide Lowenstein–Jensen mais dans ce cas, il faut obligatoirement procéder par la méthode indirecte. Les antibiogrammes faits dans un milieu liquide (Middlebrook 7H9) sont plus rapides mais plus coûteuses (Epardeau et al., 1982). Le Bactec 460-TB, très utilisé dans les années 1980, requérait un milieu liquide Middlebrook 7H12 supplémenté en acide palmitique marqué au carbone 14 (¹⁴C). La croissance bactérienne libère ainsi du ¹⁴CO₂ qui était mesuré par l'appareil. La proportion de bacilles résistants était calculée à partir de la quantité de bacilles décomptée dans les flacons témoins et ceux contenant l'antibiotique. Les résultats étaient disponibles après deux semaines.

Actuellement le système Bactec MGIT (Mycobacteria Growth Indicator Tube) est fréquemment utilisé en routine pour l'isolement des mycobactéries dans les pays riches. Celui-ci utilise un milieu Middlebrook 7H9 dans lequel un composé fluorescent est intégré. L'activité respiratoire lors de la croissance bactérienne à l'intérieur du flacon entraîne une baisse de la quantité d'oxygène dissous et permettant ainsi la détection de la fluorescence. La quantité de bactérie peut ainsi être mesurée toutes les heures par l'intensification de la fluorescence. La quantité de bactéries viables est mesurée en Colony Forming Unit par millilitre ou CFU/mL. L'antibiogramme en lecture manuelle à partir d'un milieu liquide MGIT se fait avec le coffret MGIT AST SIRE. Celui-ci se compose de 4 tubes (de 4 mL) contenant chacun un des 4 antibiotiques de première ligne et d'un tube témoin. L'absence de fluorescence dans un tube contenant un antibiotique au-delà de 2 jours après l'apparition d'une fluorescence dans le tube témoin est interprétée comme une sensibilité à cet antibiotique. L'apparition d'une fluorescence dans un tube avec antibiotique au cours des deux jours suivant celle du tube témoin révèle une résistance à l'antibiotique. L'antibiogramme à partir d'un milieu liquide Bactec MGIT

960 se fait avec des trousse Bactec MGIT 960 pour tester l'INH, l'ETB, la RIF, le PZA ou encore la streptomycine (STR). Ce sont des tubes de 7 mL. Le principe est le même que celui décrit précédemment à l'exception du fait que c'est l'automate qui effectue le calcul par comparaison du tube témoin au tube avec antibiotique. Les isolats cliniques à tester doivent être isolés sur milieu de culture solide ou liquide au préalable.

Mise à part les antibiogrammes en milieu liquide ou solide, une autre technique est aussi très largement utilisée pour déterminer la sensibilité aux antibiotiques d'un germe *M. tuberculosis*. Il s'agit de la détection moléculaire qui repose sur la mise en évidence des mutations portées par les gènes de résistance. Cette dernière technique sera décrite dans la partie méthodologie.

Le diagnostic rapide de la tuberculose et une prise en charge appropriée des personnes souffrant de tuberculose sont des étapes clés pour limiter la transmission et prévenir l'émergence de souches résistantes aux médicaments. Ces étapes constituent d'ailleurs les premières cibles des dispositions prises dans la lutte anti-tuberculeuse par les hôpitaux et les établissements de santé publics et privés dans le but de réduire toujours plus l'incidence et la prévalence de la tuberculose dans le monde.

1.1.5. La prévention de la tuberculose au moyen de la vaccination par le BCG (Bacille Calmette Guérin)

Mycobacterium bovis-BCG, vaccin vivant, est le résultat de plusieurs repiquages, sur des tranches de pomme de terre (230 passages), réalisés par Calmette et Guérin. A partir de 1921, cette souche atténuée fut utilisée comme vaccin antituberculeux. Mais les recommandations relatives à la vaccination obligatoire par le BCG en France ont été modifiées en 2007. La faible incidence de la tuberculose a permis sa limitation aux personnes à risque, tels que les nourrissons résidant dans les régions à risque élevé, telles l'Île-de-France et la Guyane française, et depuis 2012 Mayotte.

1.2. Épidémiologie descriptive de la tuberculose

La tuberculose est une maladie relevant de la santé publique par son caractère potentiellement épidémique. Elle figure sur la liste des maladies à

déclaration obligatoire depuis 1964 en France, de ce fait, l'émergence d'éventuelles épidémies est fortement surveillée d'autant plus que l'incidence et la prévalence de l'infection dans certains pays sont préoccupantes.

1.2.1. Incidence, prévalence et mortalité tuberculeuse dans le monde

L'Organisation Mondiale de la Santé (OMS) estima l'incidence mondiale de la tuberculose maladie à 9,4 millions de nouveaux cas en 2009, ce qui revient à 137 cas pour 100 000 habitants. En 2009, on a estimé le nombre de décès dus à la tuberculose à 1,3 millions. Les chiffres n'ont donc pas baissé depuis la mise en place de la stratégie *Halte à la Tuberculose* lancée par l'OMS en 2006, où l'incidence était de 139 cas pour 100 000 habitants soit 9,2 millions de nouveaux cas et la prévalence de la tuberculose s'élevait à 14,4 millions de cas. Celle-ci fut encore évaluée en 2009 à 14 millions de cas. La tendance est restée la même pour ces deux années, puisque les régions qui comptaient le plus de cas étaient l'Asie du Sud-Est (avec l'Inde et la Chine) et l'Afrique notamment l'Afrique subsaharienne. En effet 35% des nouveaux cas de tuberculose enregistrés dans le monde provenaient de l'Asie du Sud-Est et 30% d'Afrique.

Région de l'OMS	Incidence ¹		Prévalence ²		Mortalité (hors VIH)		
	Nombre en milliers	% du total mondial	Pour 100 000 habitants	Nombre en milliers	Pour 100 000 habitants	Nombre en milliers	Pour 100 000 habitants
Afrique	2800	30,00%	340	3900	450	430	50
Amériques	270	2,90%	29	350	37	20	2,1
Méditerranée orientale	660	7,10%	110	1000	180	99	18
Europe	420	4,50%	47	560	63	62	7
Asie du Sud-Est	3300	35,00%	180	4900	280	480	27
Pacifique occidental	1900	21,00%	110	2900	160	240	13
Ensemble du monde	9400	100,00%	140	14000	164	1300	19

Tableau 2. Estimation de l'incidence, de la prévalence et de la mortalité en 2009 dans les différentes régions de l'OMS. Ce tableau provient du site : <http://www.who.int/mediacentre/factsheets/fs104/fr/>.

¹Incidence: nouveaux cas survenant pendant une période déterminée.

²Prévalence: nombre de cas existant dans la population à un moment déterminé.

Sans surprise, ces deux régions présentaient également le plus grand nombre de décès dus à la tuberculose (hors VIH) avec 27 décès pour 100 000 habitants en Asie

du Sud-Est et 50 décès pour 100 000 habitants en Afrique pour l'année 2009. Cet état des lieux de la tuberculose fut signalé dans le rapport *Global tuberculosis control 2010* par l'OMS (tableau 2).

La France, pays à faible incidence, présentait en 2009, 5 276 cas de tuberculose déclarée, ce qui revient à 8,2 cas pour 100 000 habitants. L'incidence de la tuberculose en France est relativement stable depuis 1990, Cependant celle-ci masque une disparité régionale importante. De plus, la tuberculose est principalement retrouvée dans les situations précaires (cas des SDF) ou encore chez les immigrés dont le pays d'origine présente une forte incidence pour la tuberculose (cas de la plupart des pays africains et asiatiques).

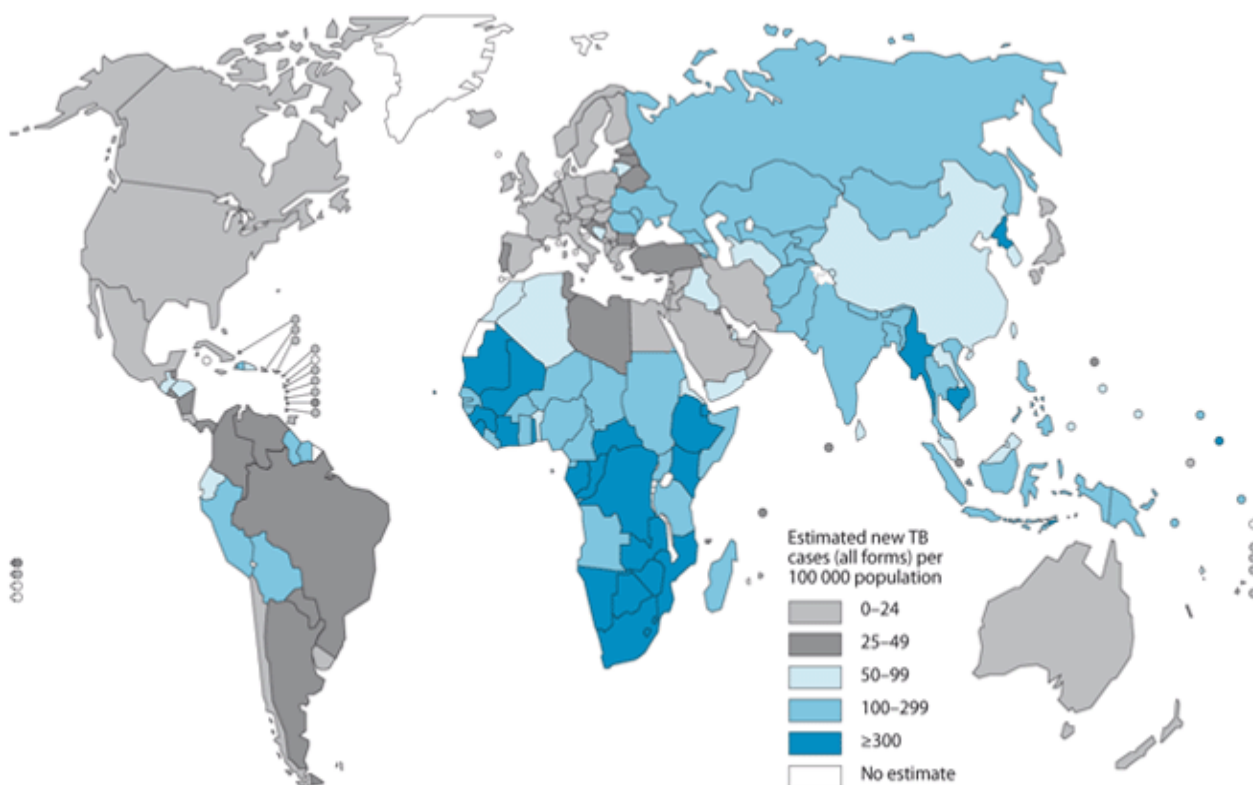


Figure 2. Estimation de l'incidence dans les différents pays du monde 2009. Carte tirée du rapport *Global tuberculosis control* : WHO report 2010, (http://www.who.int/tb/publications/global_report/2010/en/index.html).

1.2.2. L'Afrique et Asie du Sud-Est, zones à forte incidence selon l'OMS

En 2007, d'après l'OMS les cinq pays les plus concernés par des cas de tuberculose étaient l'Inde (2 millions), la Chine (1,3 millions), l'Indonésie (530 000),

l'Afrique du Sud (460 000) et le Nigéria (460 000). L'inégalité de la répartition des cas de tuberculose dans le monde est chose évidente. Les trois principaux facteurs responsables de la persistance de cette forte prévalence sont la pauvreté, la croissance démographique et l'infection par le VIH, propice au développement de la maladie. Cependant, la négligence des autorités sanitaires n'est pas en reste puisque la tuberculose n'a pas longtemps été considérée comme une priorité sanitaire nationale. L'insuffisance des dispositifs mis en place et les lacunes dans les programmes nationaux pour la couverture sanitaire et le développement des laboratoires ont contribué à la non détection des personnes malades et contagieuses ce qui a laissé le champ libre à la transmission. Il est difficile d'entrevoir un renversement de la situation surtout dans les pays africains. En effet, la pauvreté, la croissance démographique, l'exode rural, les guerres et conflits et l'épidémie de VIH/SIDA sont autant d'entraves au succès des programmes de lutte pour éradiquer la tuberculose. Entre 1990 et 2007, l'incidence de la tuberculose fut multipliée par deux en Afrique noire.

1.3. Les préoccupations tuberculeuses actuelles

1.3.1. Une tuberculose pharmaco-résistante

La pharmaco-résistance aux médicaments antituberculeux est un problème majeur de santé publique, elle apparaît quand les antibiotiques sont mal utilisés chez des patients dont la tuberculose est sensible.

La tuberculose multirésistante met en cause des mycobactéries résistantes à au moins l'isoniazide ou la rifampicine, les deux antibiotiques les plus efficaces contre la tuberculose. La résistance acquise de ces souches est le résultat de l'administration de schémas thérapeutiques inadaptés par les agents de santé ou le mauvais suivi des patients sur la prise de l'intégralité de leur traitement. La chimiothérapie des patients atteints de tuberculose MDR devient alors plus lourde par sa durée (2 ans), mais aucun coût et des effets secondaires accrus. D'après l'OMS, en 2009 la proportion des cas de tuberculose MDR fut estimée à 5% de l'ensemble des nouveaux cas de tuberculose, ce qui correspond à 440 000 nouveaux cas dans le monde. L'Inde, la Chine, la Fédération Russe, l'Afrique du Sud et le Bangladesh sont les pays les plus touchés (WHO, 2010). Les contraintes liées à la thérapie requise dans ces cas, compliquent l'accès aux soins dans ces pays ce

qui accentue la menace à la santé de ces populations. Parmi les nouveaux cas, on compte 150 000 décès et seuls 7% des cas de tuberculose résistante sont diagnostiqués et traités. La moitié des cas de résistance dans le monde est enregistrée en Chine et en Inde. La maîtrise de la tuberculose pharmaco-résistante constitue un réel défi puisqu'elle réclame un effort politique pour combler l'insuffisance des programmes de surveillance et de prévention de la tuberculose résistante en mobilisant les ressources financières pour développer les services de laboratoire de qualité et l'approvisionnement en médicaments antituberculeux.

1.3.2. Le Virus de l'Immunodéficience Humaine (VIH) et le Bacille de Koch (BK) : un tandem meurtrier

Le virus de l'immunodéficience humaine est un terrain propice au développement de la tuberculose-maladie. En 2009, 1,8 millions de personnes sidéennes sont décédés à cause de la tuberculose. Un quart des décès de personnes tuberculeuses était lié à l'infection par le VIH. Plus de la moitié des individus co-infectés par le VIH et la tuberculose se trouve en Afrique subsaharienne qui regroupe à elle seule les deux tiers des personnes atteintes par le VIH. Les épidémies de VIH et de tuberculose se renforcent mutuellement. En effet, une personne souffrant du VIH peut connaître une réactivation endogène d'une tuberculose préexistante, peut faire une réinfection ou rechute, dans ces deux derniers cas cette personne va participer à la transmission de bacilles de la tuberculose dans la population. Si les efforts requis pour arriver à une baisse de l'incidence de la tuberculose dans les pays en voie de développement réclament un investissement financier important et le renforcement des structures de santé, la concertation avec les programmes nationaux de lutte anti-VIH s'ajoute à la difficulté de la tâche.

1.4. Les programmes de lutte antituberculeuse

1.4.1. La stratégie DOTS (Directly Observed Treatment Short-Course)

Depuis 1993, la tuberculose est considérée comme une urgence sanitaire mondiale. Pour prévenir et garantir la guérison de patients tuberculeux, l'OMS a mis en place la stratégie DOTS en 1995. Cette procédure s'appuie sur 5 axes :

- Engagement des pouvoirs politiques au renforcement de tous les services,

- Dépistage de la tuberculose par un examen microscopique direct chez les sujets susceptibles d'être infectés,
- Utilisation d'un schéma thérapeutique standard, s'étalant sur 6 à 8 mois pour tous les cas diagnostiqués sous surveillance directe d'un professionnel de santé,
- Approvisionnement sans coupure en antituberculeux (les plus efficaces),
- Notification pour suivre la réponse au traitement et évaluer l'efficacité du programme.

Entre 1995 et 2008 la stratégie DOTS a permis le traitement de 36 millions de patients. Le taux de décès parmi les malades est passé de 7,6% à 4%. Depuis 2004 on a observé une régression continue de l'incidence de la tuberculose dans le monde. Cette baisse n'a malheureusement pas été observée dans les pays africains. D'ailleurs cette régression (inférieure à 1% par an) est très lente. De plus avec la croissance démographique le nombre absolu de cas croît.

1.4.2. La stratégie Halte à la tuberculose

La stratégie *Halte à la Tuberculose* fut lancée par l'OMS en 2006. Celle-ci s'appuie sur le succès de la stratégie DOTS. L'accent est mis sur la co-infection VIH/tuberculose et la prévention de l'émergence des souches résistantes. Cette nouvelle stratégie s'est donné pour date butoir l'année 2015 pour éradiquer la tuberculose. Ce plan mondial se compose de six points importants :

- Assurer l'expansion de la stratégie DOTS,
- Combattre la co-infection VIH/tuberculose et lutter contre la tuberculose résistante,
- Améliorer les systèmes de santé,
- Mettre l'accent sur l'éducation des patients,
- Impliquer la collectivité dans la lutte,
- Elargir la place occupée par la recherche.

L'état des lieux fait en 2009 par l'OMS (voir la partie : Incidence, prévalence et mortalité tuberculeuse dans le monde) donne une idée de l'impact de ces nouveaux

dispositifs sur le terrain. Les différents programmes de lutte lancés pour contrer la tuberculose donnent des résultats mitigés faisant une césure entre les pays développés et ceux en voie de développement. Si les stratégies de lutte ont été standardisées, alors les moyens économiques et le contexte politique et social sont plus ou moins favorables à leurs applications d'un pays à un autre.

2. Description des bacilles du genre *Mycobacterium*

Les mycobactéries sont des bacilles aérobies de 2 à 5 μm , légèrement incurvés, appartenant au genre *Mycobacterium*. Les mycobactéries sont classées parmi les bactéries à Gram positif quoique leur paroi présente une structure particulière qui leur confèrent leur propriété d'acido-alcool-résistance.

2.1. Une paroi cellulaire caractéristique

2.1.1. Généralités sur la paroi cellulaire des bactéries à Gram +

Les bactéries *Gram positif* sont des bactéries qui répondent à la coloration de Gram de par la structure de leur paroi cellulaire. Celui-ci se compose généralement de trois parties :

- la couche externe composée de peptidoglycane,
- l'espace périplasmique (ou périplasme) occupant une position intermédiaire,
- la membrane plasmique (ou membrane interne), constituant la partie interne de la paroi cellulaire.

La **couche de peptidoglycane**, très épaisse, assure le maintien de la topologie cellulaire. Elle joue aussi le rôle de barrière protectrice contre d'éventuelles contraintes mécaniques. Sa structure chimique peut être décrite par une suite alternée de couches de polymère de N-acétyl-glucosamine (NAG) et d'acide N-acétylmuramique (NAM). Des chaînes d'acides lipotéichoïques (ALT), enchâssées dans les acides gras de la membrane plasmique et traversent la paroi de part en part. Les polymères d'acides téichoïques quant à eux, assurent l'attache des peptidoglycanes à la membrane, et renforcent ainsi la stabilité cellulaire.

Le périplasme, espace réservé à beaucoup de réaction biochimique, concentre des nutriments, des enzymes où encore des protons rejetés lors de la respiration cellulaire.

Enfin, la **membrane interne** est une bicouche lipidique contenant des peptidoglycanes, de protéines membranaires et est comparable à celle des bactéries à Gram +. Cependant, cette couche ne contient jamais d'acides téchoïques. Cette couche est fortement impliquée dans la croissance et le métabolisme bactérien.

La coloration de Gram, dont le protocole fut établi en 1884 par le bactériologiste, Hans Christian Gram, permet la mise en évidence rapide des propriétés de la paroi bactérienne. Elle suit les étapes suivantes :

- Coloration au violet *de gentiane* (se fixe aux constituants du cytoplasme) puis au *Iugol* (Mordançage),
- Décoloration à l'alcool.

Les bactéries qui conservent la coloration violette, sont qualifiées de bactéries à Gram positif car possèdent une couche épaisse de peptidoglycanes reste imperméable à l'alcool. Les bactéries roses attestent d'une paroi de peptidoglycanes fine avec membrane externe, ce sont les bactéries Gram négatif.

2.1.2. Paroi cellulaire des mycobactéries

La paroi cellulaire mycobactérienne se compose de trois couches distinctes. La première couche dite basale a une épaisseur comprise entre 3 et 13 nanomètres (nm) (figure 3). Elle est composée de plusieurs couches de peptidoglycanes où sont enchâssés des arabino-galactanes (polymères constitués d'arabinose et de galactose). La seconde couche mesure 8 nm d'épaisseur, elle est constituée de mycolate d'arabino-galactane. Les acides mycoliques ont une chaîne principale (en C50-C60) et plusieurs chaînes ramifiées (en C24-C26). Ces grosses chaînes d'acide gras représentent environ 8% du poids sec des bacilles tuberculeux (Asselineau et Lederer, 1951). La troisième couche (membrane externe), se compose de lipo-arabinomannane, de mycolate de tréhalose, de sulfolipides et de phospholipides. Le mycolate de tréhalose encore appelé *cord factor* est retrouvé uniquement chez les souches virulentes. Le *cord factor* permet aux bacilles tuberculeux de former de longues cordes.

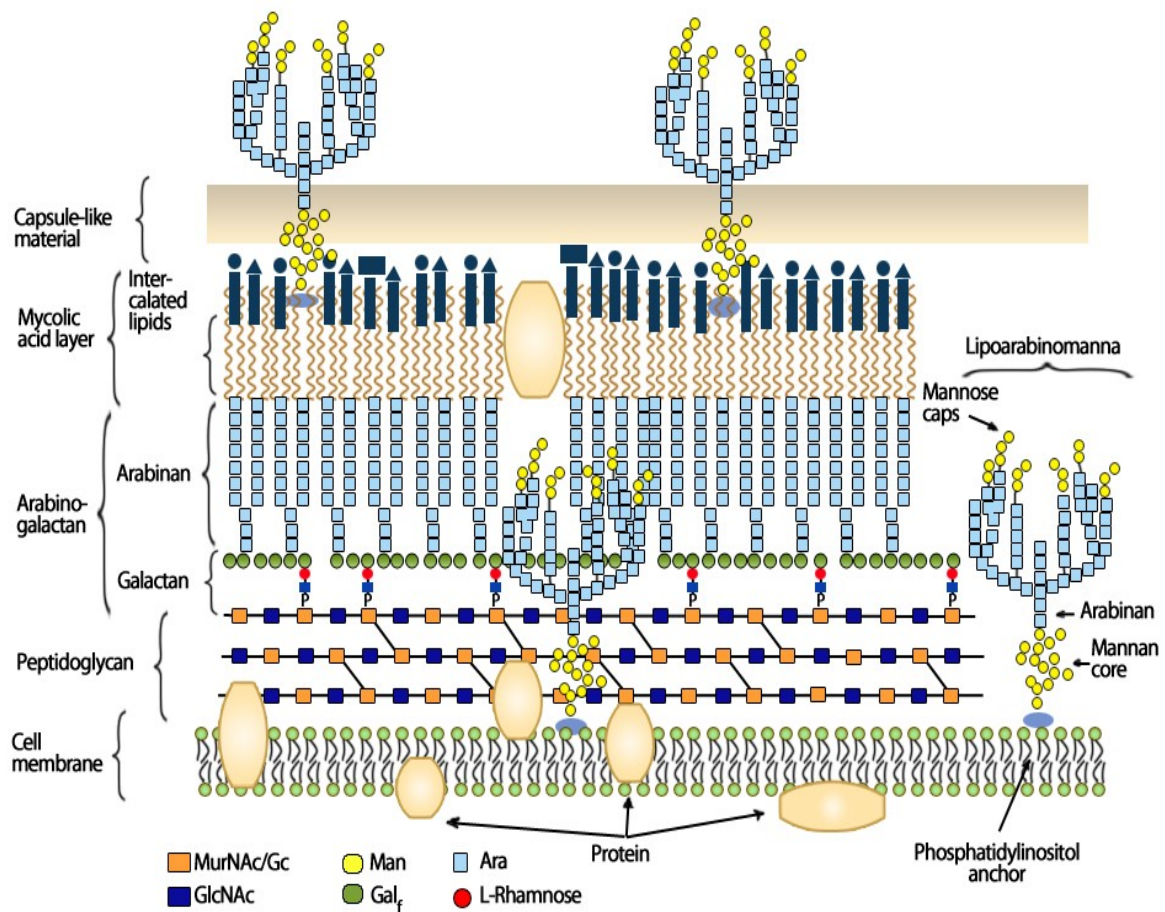


Figure 3. Composition membranaire de l'enveloppe mycobactérienne. (Image tirée du site web : <http://2009.igem.org/wiki/images/1/17/MbMyco.png>)

2.1.3. Paroi cellulaire et virulence mycobactérienne

La paroi cellulaire renferme beaucoup de facteurs de virulence, qui ont fait l'objet de nombreuses études. Certains facteurs réduisent l'immunité à l'intérieur des cellules et participe à la persistance de ces microorganismes dans l'hôte. Les lipoarabinomannanes (LAM), glycolipides présents dans la paroi, sont responsables de l'internalisation de la bactérie dans le macrophage et agirait sur l'immunité de l'hôte en bloquant la maturation phagosomale, processus durant laquelle le phagosome interagit avec les endosomes et les lysosomes pour la destruction des bactéries (Rajni et al., 2011). Le *cord factor* joue également un rôle dans l'atténuation de la réponse immunitaire en inhibant la migration des granulocytes neutrophiles qui sont des acteurs de la phagocytose. Les sulfolipides, quant à eux, bloquent la fusion des phagosomes et des lysosomes (Goren et al., 1976) et parviennent à moduler la

réponse oxydative et la sécrétion des cytokines par les neutrophiles et monocytes (Brozna et al., 1991 ; Pabst et al., 1988).

2.2. Comportement des mycobactéries au sein de l'hôte sensible

Les bacilles de Koch (BK) s'introduisent dans un hôte, généralement dans les voies respiratoires (tuberculose pulmonaire) par inhalation de gouttelettes de Pflügge, chargées en BK et atteignent les alvéoles pulmonaires. Une fois arrivées dans les poumons, les bactéries ont quatre issues. (i) La réponse immunitaire peut être radicale en les éliminant toutes, ainsi les personnes concernées ne développeront jamais une tuberculose clinique. Il arrive aussi, mais de façon moins fréquente, qu'après l'infection les (ii) bacilles se multiplient pour déclencher une tuberculose primaire. (iii) L'infection peut être contenue en une tuberculose latente, dans ce troisième cas, elle peut être détectée par un test à la tuberculine. La tuberculose latente peut persister indéfiniment ou (iv) évoluer en une forme active lors d'une seconde exposition aux bactéries ou lorsque l'hôte se retrouve dans un contexte de vulnérabilité favorable à ce changement (infection par le VIH, malnutrition...).

2.2.1. Le granulome, un agrégat de macrophages pour l'isolation du site de l'infection

Au niveau des alvéoles pulmonaires se développent des granulomes dans lesquels interviennent les effecteurs de la défense immunitaire pour contenir et éradiquer l'infection. Les bactéries sont phagocytées par les macrophages. Ceux-ci, tout comme les cellules dendritiques, présentent des antigènes bacillaires aux lymphocytes T.

Cependant, chez les cellules dendritiques, un compartiment intracellulaire présente des conditions bactériostatiques qui empêchent la croissance intracellulaire de *M. tuberculosis*. La liaison de *M. tuberculosis* aux récepteurs des cellules dendritiques de type lectine (DC-SIGN) permet l'internalisation de la bactérie. Les composés sucrés de la paroi mycobactérienne se fixent aux récepteurs dendritiques. Cette interaction favoriserait une tuberculose latente.

L'interaction macrophage-lymphocyte, quant à elle, active les macrophages alvéolaires dans la production des molécules réactives oxygénées et d'espèces

réactives oxygénées et azotées pour augmenter leur action bactéricide. Les lymphocytes les plus sollicités dans la réponse immunitaire à médiation cellulaire sont les lymphocytes CD4⁺ et les lymphocytes CD8⁺. Les lymphocyte CD4⁺ se répartissent en deux groupes les Th1 et les Th2 (Delespesse et al., 1997 ; Romagnani et al., 1996). Les cellules Th1 sécrètent essentiellement les cytokines IFN- γ et IL-2, alors que les Th2 produisent les IL-4, IL-5 et IL-10, Les CD4⁺, par l'action des cytokines sécrétés, intensifient la réponse immunitaire en activant les cellules effectrices et en recrutant des cellules supplémentaires. Chaque cytokine a une action particulière, par exemple le « *tumor necrosis factor α* » (TNF α) active l'endothélium pour induire la fixation des macrophages et leur sortie des vaisseaux sanguins pour leur acheminement sur le site de l'infection.

Les granulomes sont aussi le siège d'intenses réactions de défense que les mycobactéries arrivent parfois à déjouer grâce à leurs facteurs de virulence qui menacent l'intégrité de ces granulomes. Les chémokines (cytokines) stimulent l'afflux des macrophages sur le point d'infection. Les IFN- γ activent les macrophages et détruisent les bactéries ingérées. On peut conclure en affirmant que la production de cytokines par les CD4⁺ stimule d'autres cellules qui renforcent la défense immunitaire. Il en découle une intensification de la phagocytose et de la destruction des mycobactéries dans les macrophages. La réaction cellulaire immune d'hypersensibilité retardée (type IV) produit des lésions aboutissant à la destruction tissulaire et s'accompagne de la formation de macrophages épithélioïdes et de cellules géantes de Langhans (fusion de plusieurs macrophages). Les granulomes mycobactériens, lésions contenant les cellules épithélioïdes et de Langhans, sont à l'origine des nécroses caséuses caractéristique de la tuberculose. Cet environnement hostile, à l'intérieur des granulomes, neutralise les bactéries.

Les lymphocytes CD8⁺ sont directement cytotoxiques pour les cellules cibles. Ils reconnaissent les antigènes traités dans le cytosol et présentés dans un contexte CMH de classe I (molécule retrouvée à la surface de presque toutes les cellules nucléées). Les CD8⁺ inhibent la croissance des mycobactéries en lysant les cellules infectées avec le soutien de protéines cytolitiques, granulysines et perforines (Schluger and Rom, 1998).

2.2.2. Tuberculose pulmonaire et extra pulmonaire

Dans 85% des cas, la tuberculose se manifeste sous une forme pulmonaire mais elle peut aussi exister sous d'autres formes. Si dans la majorité des cas, l'infection se limite au foyer initial et au ganglion satellite, il peut survenir des problèmes avec la réponse immunitaire qui entraîne une dissémination de la mycobactérie. Cette prolifération a lieu par voie sanguine (ou voie hématogène) pour s'étendre à d'autres organes et tissus, on parle ici de tuberculose miliaire.

Tuberculose urogénitale. Dans ce cas, l'infection atteint le tractus urogénital par voie hématogène en partance des poumons touchés par *Mycobacterium tuberculosis*. Les lésions urogénitales sont souvent postérieures aux symptômes déclarés lors de la tuberculose active pulmonaire. Ces lésions peuvent être unilatérales ou bilatérales. L'infection siège dans la médullaire rénale où se développent des granulomes épithélioïdes accompagnés de nécroses caséeuses qui détruisent les tissus. Dans un quart des cas, on arrive à une calcification du rein. L'infection peut se propager aux uretères et à la vessie avec des lésions granulomateuses. Dans trois quarts des cas, la tuberculose urogénitale est associée à une atteinte rénale. Si l'infection se limite au rein, elle peut être asymptomatique et conduire ainsi silencieusement à la destruction progressive du rein. Malheureusement, les douleurs et l'hématurie rénales qui peuvent donner l'alerte, restent rares d'où le fait que le diagnostic soit souvent tardif (Wafra et Michel, 2005).

Tuberculose osseuse. Elle peut toucher n'importe quelle partie du squelette. Certes, les localisations osseuses préférentielles sont la colonne vertébrale (appelée le mal de Pott), la hanche et le genou (plus fréquent).

Tuberculose digestive. L'atteinte intestinale s'explique le plus souvent par les mycobactéries ingérées.

Tuberculose ganglionnaire. Après un trouble immunitaire, la tuberculose pulmonaire peut se compliquer en une tuberculose ganglionnaire par voie hématogène. Elle se caractérise par le développement de ganglions anormaux surtout au niveau du cou. Ceux-ci deviennent durs mais restent non douloureux.

Méningite tuberculeuse. Elle désigne la formation de foyers méningés. La formation de granulomes dans les méninges résulte de la migration hématogène des bacilles suite à une primo-infection non traitée. On retrouve donc des bacilles tuberculeux dans le liquide céphalo-rachidien.

Pour conclure cette partie, on peut affirmer que les infections par mycobactéries peuvent siéger dans différents organes ou tissus biologiques. Il existe plusieurs espèces de mycobactéries et chacune d'elles a des conséquences caractéristiques quant à la localisation de l'infection tuberculeuse.

2.3. La classification des mycobactéries et description des diverses espèces mycobactériennes

La classification tend à donner une représentation claire du vivant en le scindant en plusieurs catégories. Avant d'arriver à la classification phylogénétique, largement considérée, plusieurs méthodes de classification se sont tour à tour imposées dans la communauté scientifique. Les premières classifications, classification populaire et classification primitive étaient toutes deux très sommaires. Elles se basaient sur l'observation de caractères morphologiques communs mêlés à des croyances populaires. L'accumulation croissante des connaissances biologiques a mené sans cesse à reconsidérer des critères et/ou à les conforter par des nouveaux, ce qui a entraîné de façon inévitable à une évolution dans la méthodologie employée pour retranscrire l'ordre du vivant et finalement à des méthodes de classification plus robustes que les premières utilisées.

2.3.1. Les principales méthodes de classification biologique

La **classification traditionnelle** est largement basée sur les travaux du naturaliste suédois Carl von Linné. Initialement, Linné avait divisé le vivant en 3 règnes : les règnes animal, végétal et minéral. Le nombre de règnes s'est accru pour atteindre 6 avec les bactéries, les archées, les protistes, les champignons, les végétaux et les animaux. Chaque règne est hiérarchisé en 6 catégories comme suit : *l'embranchement (division ou encore phylum) → la classe → l'ordre → la famille → le genre → l'espèce*. Cette hiérarchisation permet une caractérisation graduelle basée sur des critères biologiques, phénotypiques et physiologiques. Par exemple, la systématique mycobactérienne se décline comme suit, règne : *Bacteria* → classe : *Actinobacteria* → sous-classe : *Actinobacteridae* → ordre : *Actinomycetales* → sous-ordre : *Corynebacterineae* → famille : *Mycobacteriaceae* → genre : *Mycobacterium*.

La **classification phylogénétique** remplaça la classification traditionnelle par l'adhésion à la théorie de l'évolution qui introduit l'idée d'une ascendance commune parmi les organismes. Cette classification pose ses fondements sur la cladistique qui reconnaît le clade comme unité de classe. Un clade regroupe un ancêtre avec tous ses descendants (on parle aussi de groupe monophylétique). La classification phylogénétique considère donc les relations de parenté qui se sont tissées durant la différenciation successive des espèces au fil des générations. Elle obéit donc à un schéma évolutif. Le résultat graphique de cette méthodologie est un cladogramme qui est un graphe où les groupes monophylétiques s'emboîtent de façon hiérarchique puisque des rangs supérieurs aux rangs inférieurs, il y a une croissance du nombre de caractères dérivés en commun.

Cette classification regroupe l'ensemble des êtres vivants dans 3 clades : les eubactéries (organismes unicellulaires à paroi faite de peptidoglycane), les archées (organismes unicellulaires à paroi lipidique), et les eucaryotes (organismes uni/multicellulaires dont le matériel génétique se trouve dans un noyau). Les procaryotes regroupent les archées et les eubactéries.

Les travaux de phylogénie moléculaire de Carl Woese (nomenclature de 1990) regroupèrent sous le terme de procaryotes, les bactéries (précédemment appelées eubactéries) et les archées.

2.3.2. La notion d'espèce bactérienne

Dans toutes les méthodes de classification la notion d'espèce occupe une place importante. Chez les procaryotes, la définition d'espèce n'a pas été facile à établir puisqu'ils ne se multiplient pas par reproduction sexuée. En effet, il est communément admis qu'une espèce biologique regroupe des individus partageant un ensemble de caractères, pouvant non seulement se reproduire (sexuellement) entre eux mais également engendrer des progénitures fertiles. Cette définition ne pouvant s'étendre aux procaryotes, il a fallu considérer d'autres critères pour décrire une espèce bactérienne. En 1987, un comité du nom de «International Committee on Systematic Bacteriology (ICSB)» donna une description phylogénétique d'une espèce bactérienne. Cette définition est basée sur l'homologie ADN-ADN de plusieurs milliers de souches provenant de 2000 d'espèces différentes. Sont

considérées comme faisant partie de la même espèce, les souches ayant en commun une valeur d'hybridation d'au moins 70% et une différence de la température de fusion T_m (melting temperature) inférieure ou égale à 5°C. Cette méthodologie basée sur l'hybridation d'ADN atteste de la proximité des souches (Wayne, 1988).

2.3.3. Diversité des espèces mycobactériennes et apport des marqueurs moléculaires

2.3.3.1. Généralités sur le chromosome mycobactérien

Les mycobactéries présentent un chromosome unique circulaire. Chez *Mycobacterium tuberculosis* H37Rv, souche de référence, il y a 4,41 millions de paires de bases, 65,6% de GC, 4062 gènes et 8 pseudo-gènes. H37Rv fut isolée en 1934 d'un poumon humain. Depuis, il a largement été utilisé à travers le monde dans le domaine biomédical. Sa virulence est maintenue, elle est sensible aux antibiotiques et réceptive aux manipulations génétiques.

Les mycobactéries se multiplient par reproduction clonale. Elles sont soumises à une reproduction à l'identique par le clivage d'une cellule en deux cellules. Le temps de génération pour *Mycobacterium tuberculosis* est de 24 heures (il est important ici de signaler que ce délai varie en fonction du milieu). Malgré une reproduction asexuée par scission binaire, le chromosome mycobactérien connaît des variations génétiques ou mutations qui peuvent être maintenues dans le temps, si ceux-ci ne confèrent pas aux individus qui les portent un phénotype défavorable à leur environnement.

2.3.3.2. Le polymorphisme génétique : support de la différenciation des souches MTC au cours du temps

Une mutation génétique désigne toute modification survenant sur la séquence d'ADN. Il existe plusieurs types de mutations : les mutations par substitution (remplacement d'un nucléotide par un autre), les mutations par insertion ou délétion d'un nucléotide.

Si la mutation a lieu dans un cadre ouvert de lecture, elle peut avoir un impact plus ou moins préjudiciable à la fonction de la protéine qui sera traduite. En effet, il existe plusieurs types de polymorphisme qui peuvent entraîner des variations entre

individus de même espèce. On a d'abord pensé que *Mycobacterium tuberculosis* présentait peu de diversité génétique (Kapur et al., 1994 ; Sreevatsan et al., 1997c), mais cette affirmation fut mise en cause suite aux travaux de Fleischmann et col. (2002) (figure 4).

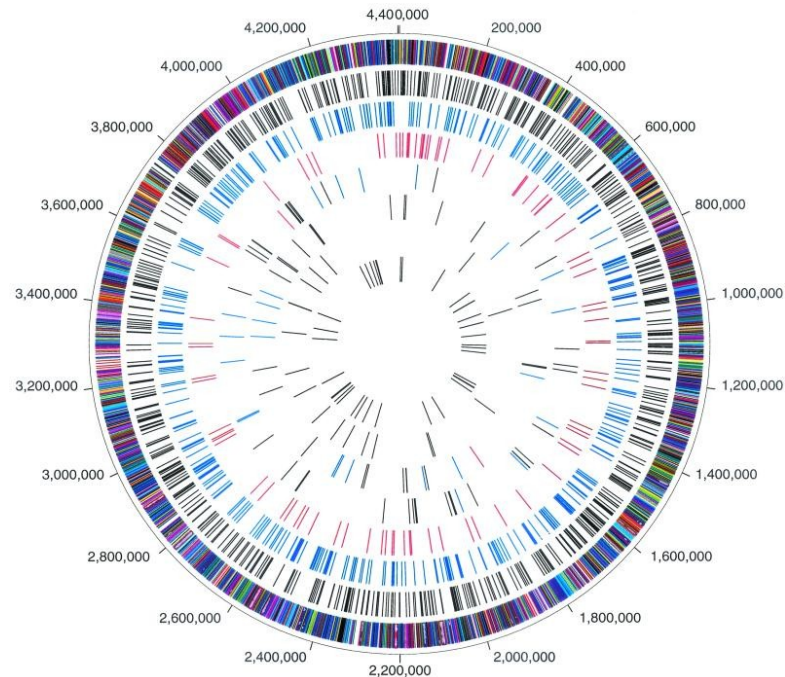


Figure 4. Représentation circulaire du chromosome de *M. tuberculosis* indiquant l'emplacement de chaque protéine-région codante prédite ainsi que les fonctions sélectionnées différentes entre les souches CDC1551 et H37Rv. Le cercle concentrique externe montre les régions codantes pour la protéine prédite sur les deux brins. Le 2^{ème} cercle concentrique montre l'emplacement des substitutions non synonymes (noir). Le 3^{ème} cercle concentrique montre l'emplacement de substitutions synonymes (bleu). Le 4^{ème} cercle concentrique montre l'emplacement des substitutions dans les régions non codantes (rouge). Le 5^{ème} cercle concentrique montre l'emplacement des insertions dans la souche CDC1551 et l'emplacement du phage phiRv1 (rouge). Le 6^{ème} cercle concentrique montre l'emplacement des insertions dans la souche H37Rv et l'emplacement du phage phiRv1 (rouge). Le 7^{ème} cercle concentrique montre l'emplacement des éléments d'insertion IS6110 dans les souches CDC1551 (bleu) et H37Rv (rouge). Le 8^{ème} cercle concentrique montre l'emplacement des ARNt (bleu) et d'ARNr (rouge). Image tirée de l'article de Fleischmann et al. (2002).

2.3.3.2.1. Les régions de différence chez *M. tuberculosis*, support des « Large Sequence Polymorphism » (LSP)

La comparaison du génome de H37Rv à celui de CDC1551, a permis de mettre en évidence 37 insertions correspondant à des régions de longueur supérieure à 10 pb présentes chez H37Rv et absentes de chez CDC1551. Ces insertions concernent en grande partie (26 cas) des cadres ouverts de lecture (en anglais : open reading frame, ou ORF) et aussi des régions intergéniques (11 cas).

Ces insertions concernent non seulement de courtes séquences répétées en tandem à l'extrémité 3' ou 5' des ORFs, mais aussi des ORFs en entier. A l'inverse, il y a des insertions chez CDC1551 que l'on ne retrouve pas chez H37Rv. Près de la moitié de ces ORFs polymorphes codent pour des protéines appartenant à la famille des PE (proline acide glutamique) et PE_PGRS (proline acide glutamique - séquence polymorphique riche en GC). Le prophage phiRv1, que l'on retrouve dans la région RD3 chez *Mycobacterium bovis* (Mahairas et al., 1996), est localisé dans une famille de répétition appelée REP13E12 (Gordon et al., 1999), mais sa présence est variable. Divers RDs (Region of Difference) qui font partie des Large Sequence Polymorphism (LSP) furent identifiées au sein du complexe *Mycobacterium tuberculosis* (MTC). Ces longues délétions de séquence impliquent plusieurs ORFs et ont de ce fait des conséquences phénotypiques. Par exemple, l'absence de la région RD1 qui s'étend de l'ORF Rv3871 à Rv3879c jouerait un rôle important dans la pathogénicité et serait en partie responsable de l'atténuation de la virulence chez les souches BCG (Bacille Calmette Guérin) (Ganguly et al., 2008) où elle est déléetée. Ces régions de délétion sont utilisées à des fins de discrimination des familles génotypiques au sein des MTC. Par exemple, la délétion RD(Rio) (Lazzarini et al., 2007) caractérise une sous lignée majeure de la famille spoligotypique LAM (Latin American-Mediterranean), (figure 5).

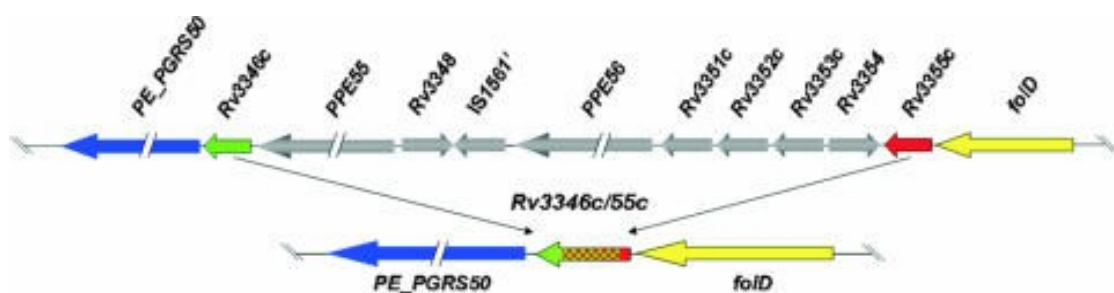


Figure 5. Schéma montrant les gènes déléetés ou altérés dans la délétion RD(Rio). La recombinaison entre Rv3346c et Rv3355c donne lieu à un nouveau gène Rv3346c/55c. La séquence génomique entre ces deux gènes déléetés aboutit à la suppression de 26 314 kb (Lazzarini et al., 2007).

Un autre exemple très étudié de LSP, la délétion spécifique à *M. tuberculosis* appelée TbD1, c'est une région génomique contenant les gènes *mmpS6* et *mmpL6*. La présence de cette région, TbD1 chez une souche MTC atteste de son caractère ancestral (Brosch et al., 2002).

2.3.3.2.2. La séquence d'insertion IS6110

IS6110 est une séquence d'insertion (IS) du complexe *Mycobacterium tuberculosis* (MTC) appartenant à la famille des IS3. Elle fut décrite pour la première fois en 1990 par Thierry et al. (1990). Cet élément transposable compte 1360 paires de bases (pb) et présente des homologies avec IS3411 retrouvée chez *Escherichia coli* (*E. coli*). La séquence d'IS6110 porte deux cadres ouverts de lecture (ORF), codant pour deux transposases impliquées dans le déplacement du transposon d'un site à un autre. Comme toutes les séquences d'insertion (IS), IS6110 est encadrée par des séquences inversées répétées (IR) de 28 paires de bases (pb) dont 3 pb mésapariées et des répétitions directes (DR, Direct Repeat) de 3 pb. Les IRs des transposons fournissent parfois un promoteur aux ORFs et constituent la cible de la transposase active. Dans une étude réalisée par Beggs et al. en 2000, l'insertion d'un IS6110 fut mise en évidence à l'intérieure du gène *ctpD* (ATPase transporteur de cation) de *M. tuberculosis*. Cette étude montra la possibilité que l'IS6110 fournit un promoteur pour la transcription des gènes. D'ailleurs, 50% des cas d'insertion d'IS6110 se retrouve dans une région codante (Sampson et al., 2001 ; Otal et al., 2007). La présence des répétitions répétées (DRs) à l'extrémité de l'IS est le résultat d'une transposition. Par exemple, l'IS6110 insérée dans le locus DR est encadré par 3 nucléotides répétées (CCC) (Beggs et al., 2000). Chez *Mycobacterium bovis* des DRs de 3-4 pb se retrouvent également à l'extrémité des IS6110 étudiées (Otal et al., 2007). Sampson affirme que l'absence de ces 3 pb répétées au point d'insertion est le résultat de recombinaison homologue entre deux IS6110 intégrées dans le même sens (Sampson et al., 2003). Ce type de recombinaison pourrait donc conduire à la perte de la séquence génomique située entre les deux IS. Dans une étude de 2002, Fleischmann et al. ont observé que 4/16 des IS6110 présentes chez H37Rv avaient perdu les DRs de 3-4 pb. Ces quatre IS6110 sont directement adjacents à des régions délétées chez la souche CDC1551 (Fleischmann et al., 2002). En ce sens, Ho et al. ont démontré, par alignement de séquences de génomes de *Mycobacterium tuberculosis*, que sur une longueur de 20 000 pb, les régions encadrant les points d'insertion d'IS6110 sont fortement variables, puisqu'elles sont le siège de délétion associées à la présence de l'IS (Ho et al., 2000). Ces résultats ont été confirmés par Sampson et al. (2003) dans une région allant de Rv1754 à

Rv1765 qui a subi beaucoup de délétion au cours de l'histoire évolutive de MTBC (Sampson et al., 2003).

D'une façon générale, les éléments transposables participent activement aux réarrangements chromosomiques. Dans une étude réalisée sur *Actinobacterium frankia* (Bickhart et al., 2009), on a montré l'existence d'un lien direct entre le grand nombre d'éléments transposables retrouvé dans leur génome et leur haut niveau de plasticité génomique, celle-ci responsable de la grande diversité d'hôtes et d'environnements auxquels cette bactérie est adaptée. Chez *Bordetella pertussis* et *B. parapertussis*, un grand nombre d'IS induirait des délétions génomiques ainsi que des réarrangements chromosomiques (Kirillov et al., 1995 ; Brinig et al., 2006). Diverses publications ont conclu que l'IS6110 participe activement au façonnage architectural du chromosome des souches du complexe *Mycobacterium tuberculosis* ce qui serait à l'origine de variations phénotypiques (Brosch et al., 1999). Des observations sur les modifications de la transposition d'IS6110 parmi des souches *Mycobacterium bovis* ayant infectées des hommes et des animaux, amènent à la conclusion d'une évolution adaptative hôte/pathogène liée aux séquences d'insertion (Otal et al., 2007). Plusieurs études réalisées depuis une décennie soulignent l'implication de l'IS6110 dans la plasticité génomique des bactéries appartenant au complexe *Mycobacterium tuberculosis* (Warren et al., 2000). De plus, *M. tuberculosis* présente des loci préférentiels pour l'insertion de l'IS6110 nommés « *ipl* » (IS6110 Preferential Locus) et subissent des modifications génomiques importantes par transposition ou recombinaison homologue des IS6110 (Fang et al., 1996). Cette implication «supposée» de l'IS6110 dans la délétion de gènes importants rend contestable l'utilisation de cet élément transposable comme un marqueur neutre pour les études phylogénétiques.

2.3.3.2.3. Le polymorphisme d'un seul nucléotide (SNP ou single-nucleotide polymorphism)

Grâce à la génomique comparative, on a pu identifier le polymorphisme d'un seul nucléotide qui se décline en trois types :

- les SNP synonymes : sans changement de l'acide aminé impliqué,
- les SNP non synonymes : avec changement de l'acide aminé impliqué,

- les SNP régulateurs influent sur l'expression protéique.

Cependant, selon sa localisation, le SNP peut avoir des conséquences variables surtout s'il se trouve dans une région codante. Sur le serveur NCBI, 40 308 SNPs sont recensés chez *M. tuberculosis*. La comparaison des deux génomes H37Rv et CDC1551 a révélé 1 075 SNPs (Fleischmann et al., 2002). Les SNPs sont utiles à l'identification de groupes de souches phylogénétiquement distincts (Filliol et al., 2006). Par exemple, le polymorphisme de nucléotide unique portés par les deux gènes *katG* et *gyrA* a permis de diviser le complexe *M. tuberculosis* en 3 groupes (Sreevatsan et al., 1997). Le groupe 1 (PGG1, Principal Group Genetic) est défini par la combinaison des allèles *KatG* codon 463 CTG (Leu) et *gyrA* codon 95 ACC (Thr). Le 2^{ème} groupe (PGG2) présente les allèles *KatG* codon 463 CGG (Arg) et *gyrA* codon 95 ACC (Thr). Enfin, le 3^{ème} groupe (PGG3) est décrit avec les allèles *KatG* codon 463 CGG (Arg) et *gyrA* codon 95 AGC (Ser).

2.3.3.2.4. Les loci CRISPR de *M. tuberculosis* support du spoligotypage

Les loci CRISPR (clustered regularly interspaced short palindromic repeats) sont largement observés chez les bactéries et les archées (Horvath et al., 2010). Ceux-ci sont composés de répétitions palindromiques directes intercalées par des espaceurs. Par définition, ces espaceurs sont des bouts de séquences génomiques étrangères provenant du génome de phages ou de plasmides. L'ensemble du locus serait transcrit en ARNm sous la direction d'un promoteur adjacent au locus. L'addition de nouveaux espaceurs serait assurée par les protéines traduites par les gènes CAS (CRISPR-Associated Gene). Ces espaceurs jouent le rôle d'une immunité acquise, car ils permettent la reconnaissance des éléments exogènes auxquels a déjà été exposé l'organisme. Grâce à ce système de reconnaissance les phages et éléments transposables sont éliminés. Les CRISPR joueraient alors un rôle de défense chez les bactéries et les archées.

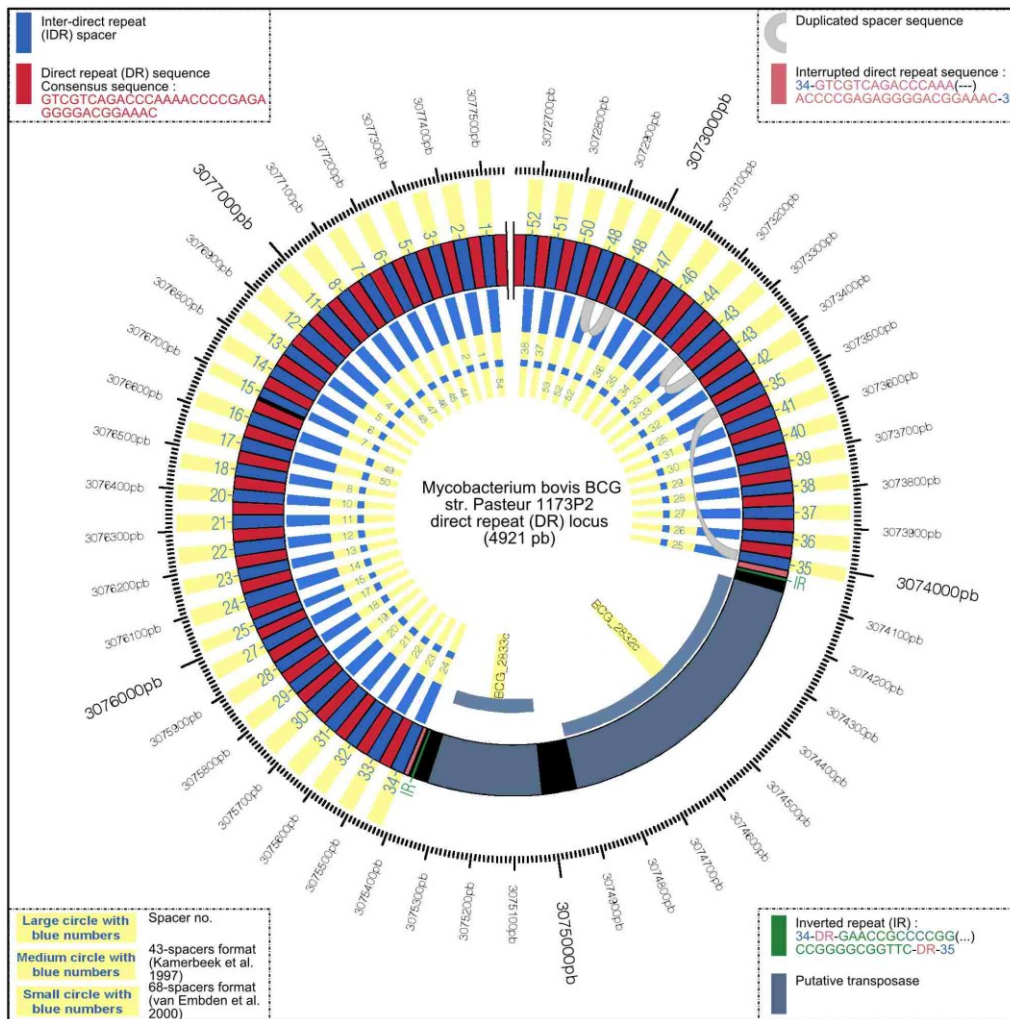


Figure 6. Représentation du locus DR de *M. bovis*-BCG, réalisée à partir des données publiées (Kamerbeek et al., 1997 ; van Embden et al., 2000) et du serveur NCBI. Cette image a été faite avec le logiciel Circos (<http://circos.ca/>).

Dans la mesure où les CRISPR sont bien conservés au cours de l'évolution, la comparaison du polymorphisme des espaceurs parmi des souches de même espèce constitue une bonne méthode de discrimination. En effet, ce principe est la base du spoligotypage très utilisé en épidémiologie et phylogénie de la tuberculose. Cette technique de génotypage a d'ailleurs permis une classification génotypique des souches de MTC (Brudey et al., 2006). Cette classification fut établie en admettant que le locus DR a accumulé au cours du temps des pertes irréversibles d'espaceurs. Le locus DR (Direct Repeat) présent chez MTC est composé de 94 espaceurs différents de 35 à 41 pb séparés par la même séquence répétées (DR) de 36 pb (van Embden et al., 2000). La figure 6 illustre le locus DR de *M. bovis*-BCG. La technique du spoligotypage sera largement détaillée dans le chapitre Matériel et Méthode.

2.3.3.2.5. Le polymorphisme du nombre d'unités répétées VNTR (Variable Number Tandem Repeats)

Les VNTR sont des minisatellites répétés en tandem dispersés dans le génome et occupant souvent une position intergénique. On peut les retrouver chez divers organismes tels *Burkholderia pseudomallei* (U'Ren et al., 2007), *Yersinia pestis* (Klevytska et al., 2001), *Streptococcus pneumoniae* (Koeck et al., 2005), etc. Les VNTR présentent un polymorphisme dans le nombre de répétitions. Chez les mycobactéries, les VNTR furent d'abord identifiés par Frothingham et al., (1998). D'autres VNTR furent par la suite observés dans le génome des mycobactéries (Smittipat et al., 2000 ; Supply et al., 2000). L'analyse de plusieurs loci VNTR (ou MLVA, Multiple Loci VNTR Analysis) est largement utilisée pour la discrimination des souches bactériennes d'une même espèce et pour l'étude des routes de transmission. Le MLVA est une technique basée sur une PCR qui évalue le nombre de répétition des différents loci VNTR. Par exemple, la base de données *MLVAbank* (<http://minisatellites.u-psud.fr/MLVAnet/>) donne accès à la diversité génotypique de divers organismes sur la base des MLVA. Depuis 2001, cette technique de génotypage est appliquée à MTC, tout d'abord en rassemblant 12 loci VNTR couramment appelés MIRU (Mycobacterial Interspersed Repetitive Units) (Supply et al., 2001). Par la suite, d'autres formats furent proposés pour améliorer le pouvoir discriminant de la méthode suivant que l'on fasse une étude épidémiologique ou phylogénétique (Supply et al., 2006).

2.3.4. Description des espèces mycobactériennes

Les marqueurs moléculaires tirant partie des régions polymorphiques du génome mycobactérien ainsi que l'identification de caractères biochimiques ont permis la caractérisation des diverses espèces de mycobactéries. On a pu ainsi diviser le genre des mycobactéries en 3 grands groupes : les mycobactéries atypiques (encore appelées mycobactéries non tuberculeuses), *Mycobacterium leprae* (agent responsable de la lèpre) et le complexe *Mycobacterium tuberculosis*.

2.3.4.1. Description des espèces appartenant au groupe des mycobactéries non tuberculeuses (MNT)

Toutes les espèces faisant partie du groupe des mycobactéries atypiques ont en commun une croissance rapide (quelques heures en milieu liquide) et se caractérisent généralement par une infection non tuberculeuse. Par opposition aux mycobactéries membres du complexe *Mycobacterium tuberculosis*, les atypiques présentent un haut niveau de résistance aux antibiotiques et, fort heureusement, une faible virulence (van Ingen et al., 2011). Ce groupe concentre une soixantaine d'espèces qui sont des pathogènes opportunistes (*M. avium*, *M. intracellulare*, *M. kansasii*, *M. marinum*, *M. malmoense*) ou saprophytes (*M. phlei*, *M. sphagni*, *M. goodnae*). Ci-après est faite une description de quelques mycobactéries atypiques.

Mycobacterium kansasii est photochromogène (production de pigment à la lumière). Elle représente 20% des infections tuberculeuses pulmonaires induites par les mycobactéries atypiques (Ohnishi et al., 2011). L'infection cutanée par *Mycobacterium kansasii* est rare et survient souvent chez des personnes immunodéprimées (60% des cas des infections cutanées dues à *Mycobacterium kansasii* concernent des immunodéprimés ; Liao et al., 2007). Son génome compte 6,4 millions de pb (Mb), 66,1% de GC et 5 961 gènes.

Mycobacterium marinum est essentiellement retrouvée dans les eaux douces et sales surtout au nord de l'Amérique. Elle infecte les grenouilles, les poissons et d'autres animaux à sang froid (Slany et al., 2012). Chez l'homme, elle cause une infection cutanée granulomateuse (parfois persistante) qui concerne surtout les extrémités des membres. *M. marinum* est photochromogène ; les colonies deviennent jaunes à la lumière. Son génome compte 6,64 millions de pb, 65,7% de GC et 5 541 gènes. On retrouve aussi chez *Mycobacterium marinum* un plasmide (pMM23) de 20 000 pb (serveur NCBI, GenBank : CP000895.1).

Le **complexe *Mycobacterium avium* (MAC)** regroupe 2 espèces distinctes : *Mycobacterium intracellulare* et *M. avium*. Ce dernier se divise en trois sous-espèces : *M. avium* sous-espèce *avium*, *M. avium* sous-espèce *paratuberculosis* et *M. avium* sous-espèce *silvaticum*. Les souches MAC se retrouvent dans les eaux stagnantes et les sols. Ces bactéries sont dites environnementales car elles sont largement répandues. Cette grande expansion est due à leur capacité de se reproduire sur une large plage de pH et de températures (jusqu'à 57 °C). L'infection

concerne les oiseaux et les personnes immunodéprimées. L'infection pulmonaire caractérise les souches MAC. Leurs colonies sont lisses et opaques ou lisses et transparentes ou encore rugueuses. Dans les deux derniers cas, on observe une croissance plus rapide. Cependant, l'ensemble des morphotypes est interchangeable. Des glycopeptidolipides spécifiques caractérisent chaque morphotype. Le génome des souches MAC représente environ 5 millions de pb et le pourcentage de GC varie entre 68% et 69%.

Le **complexe *Mycobacterium fortuitum* (MFC)** comprend deux espèces *Mycobacterium fortuitum* et *M. chelonae*. Ce sont des bactéries environnementales qui forment des colonies lisses de couleur blanchâtre ou crème. On n'observe pas de croissance au-delà de 45°C. *M. chelonae* fut identifié pour la première fois chez une tortue.

2.3.4.2. *Mycobacterium leprae* agent causal de la lèpre

Mycobacterium leprae fut découvert en 1893 par Armauer Hansen (médecin bactériologiste et dermatologue norvégien), c'est pour cela d'ailleurs que la lèpre est aussi nommée maladie de Hansen. Cette espèce est essentiellement pathogène à l'homme et se transmet principalement par contact avec les sécrétions nasales des malades non traités. Les bacilles sont en globi dans le cytoplasme macrophagique et dans les cellules de Schwann. L'infection concerne la peau et les cellules des nerfs périphériques. Le génome de cette espèce mycobactérienne ne compte que 3,27 Mb, 57,8% GC et 2 770 gènes et 1 116 pseudo-gènes. Plusieurs systèmes cataboliques ainsi que leurs voies de régulation furent supprimés, probablement à cause de délétions génomiques survenues suite à des événements de recombinaison entre séquences génomiques répétées.

2.3.4.3. Description des espèces membres du complexe *Mycobacterium tuberculosis*

Le complexe *Mycobacterium tuberculosis* (MTC) inclut *Mycobacterium africanum*, *M. caprae*, *M. bovis*, *M. microti*, *M. pinnipedii*, *M. canettii* et *Mycobacterium tuberculosis*. Dans 75% des cas, ces agents étiologiques causent une tuberculose pulmonaire (qui peut être associée à d'autres localisations), et

dans les 25% des cas restants, elle est extra-pulmonaire. Toutes ces espèces ont en commun une croissance lente.

2.3.4.3.1. Description de *Mycobacterium africanum*

M. africanum fut mis en évidence en 1968 au Sénégal par Castets (Castets et al., 1968). *M. africanum* présente un génome de 4,39 Mb, 65,6% de GC, 3 983 gènes et 103 pseudo-gènes. Par ses caractéristiques biochimiques, *Mycobacterium africanum* fut rapidement considéré comme occupant une position intermédiaire entre *M. bovis* et *M. tuberculosis*. *Mycobacterium africanum* comprend deux lignées phylogénétiques distinctes : *M. africanum* West African 1 (MWA1) and *M. africanum* West African 2 (MWA2). Cette espèce a une spécificité géographique pour l'Ouest de l'Afrique où elle représente plus de la moitié des cas (de Jong et al., 2009). Des cas furent observés ailleurs, mais souvent il s'agissait de ressortissants ouest Africains.

	^a Lignées selon les LSP		^b Lignées selon le locus DR		IS6110 RFLP	Polymorphisme de séquence					Antibiotique		Caractéristiques biochimiques	
	Nom	LSP	Nom	Espaceur absent	Nb bandes (Moy.)	KatG463	gyrA95	pncA57	oxyR 285	pks 15/1	^c TCH (1µg/mL)	^d PZA (50 - 100 µg/mL)	Production de niacine	Réduction de nitrate
<i>M. africanum</i>	West African lineage I	TbD1+ RD711- RD7- RD8- RD9- RD10-	AFRI 2	8-12, 21-24, 37-39	10	CTG (Leu)	ACC (Thr)	CAC (His)	Guanine	Δ6 pb	Sensible	Sensible	Négatif à faible	Négatif à faible
		AFRI 3	8-12, 37-39											
	West African lineage II	TbD1+ RD702+ RD7- RD8- RD9- RD10-	AFRI 1	7-9, 39	5									

Tableau 3. Caractéristiques génotypiques et biochimiques de *M. africanum*.

^a Lignées établies d'après Gagneux et Small (2007)

^b Lignées établies d'après SpolDB4 (Brudey et al., 2006)

^c TCH, hydrazide de l'acide thiophène carboxylique

^d PZA, Pyrazinamide

Cependant, MWA1 est davantage spécifique à la région du golfe de Guinée alors que MWA2 prédomine dans les régions côtières ouest Africaines (Gagneux et

al., 2006). Initialement, ces deux lignées (MWA1 et MWA2) faisaient partie de *M. africanum* type I (MAT1) par opposition à *M. africanum* type II (MAT2) qui fut ensuite reclassifié au sein de *M. tuberculosis* sous le nom de « famille Uganda » (Mostowy et al., 2004). Contrairement à *M. africanum* type I (MAT1), chez *M. africanum* type II (MAT2) la région TbD1 est absente. La présence de TbD1 et des allèles *KatG*463 CTG (Leu)/*gyrA*95 ACC (Thr) chez MAT1 en fait une espèce ancestrale du groupe PGG1 (Tableau 3). Le tableau 3 décrit les caractéristiques génotypiques et biochimiques de *M. africanum*.

2.3.4.3.2. Description de *Mycobacterium caprae*

Les chèvres constituent le réservoir le plus fréquent pour *M. caprae*. Identifié pour la première fois chez des chèvres en Espagne en 1999 (Aranaz et al., 1999), il fut nommé *Mycobacterium tuberculosis subsp. caprae*. Dans cette première étude, il fut remarqué que *M. tuberculosis subsp. caprae* ne produit pas d'acide nicotinique (niacine) et ne réduit pas le nitrate. Le polymorphisme de gène *pncA* est semblable à celui de *M. tuberculosis*. Les gènes de résistance *katG*, *gyrA* et *oxyR* ont un polymorphisme identique à celui de *M. bovis*.

^a Lignées selon les LSP		^b Lignées selon le locus DR		IS6110 RFLP	Polymorphisme de séquence					Antibiotique		Caractéristiques biochimiques	
Nom	LSP	Nom	Espaceur absent	Nb bandes (Moy.)	<i>KatG</i> 463	<i>gyrA</i> 95	<i>pncA</i> 57	<i>oxyR</i> 285	<i>pks</i> 15/1	^c TCH (1 µg/mL)	PZA (50 - 100 µg/mL) ^d	Production de niacine	Réduction de nitrate
	TbD1+ RD7- RD8- RD9- RD10- RD12 ^{bov} RD13-	BOV4-Caprae	1, 3, 16, 28, 39- 43	5	CTG (Leu)	ACC (Thr)	CAC (His)	Adénine	Δ6 pb	Sensible	Sensible +	Négatif	Négatif

Tableau 4. Caractéristiques génotypiques et biochimiques de *M. caprae*.

^a Lignées établies d'après Gagneux et Small (2007)

^b Lignées établies d'après SpolDB4 (Brudey et al., 2006)

^c TCH, hydrazide de l'acide thiophène carboxylique

^d PZA, Pyrazinamide

En 2003, *M. tuberculosis subsp. caprae* fut élevé à un statut d'espèce : *M. caprae* (Aranaz et al., 2003). En effet, la séquence du gène *gyrB* montre des substitutions

nucléotidiques que l'on ne retrouve pas ailleurs dans MTC. En considérant les LSP, on constate que *M. bovis* et *M. caprae* partagent les mêmes régions de différence (RD). Cependant, *M. bovis* a une délétion en plus (RD4). Malgré beaucoup de caractères communs avec *M. bovis*, *M. caprae* est une espèce qui lui est distincte (Aranaz et al., 2003). Le tableau 4 décrit les caractéristiques génotypiques et biochimiques de *M. caprae*.

2.3.4.3.3. Description de *Mycobacterium bovis*

M. bovis infecte principalement les bovins, cependant on le retrouve chez d'autres animaux et les humains. Ce nom fut officiellement attribué à cette espèce de mycobactérie en 1970 (Karlson et Lessel, 1970).

^a Lignées selon les LSP		^b Lignées selon le locus DR			IS6110 RFLP	Polymorphisme de séquence					Antibiotique		Caractéristiques biochimiques	
Nom	LSP	Nom	Espaceur absent	Nb bandes (Moy.)	<i>KatG463</i>	<i>gyrA95</i>	<i>pncA57</i>	<i>oxyR 285</i>	<i>pks 15/1</i>	^c TCH (1µg/mL)	PZA (50 - 100 µg/mL) ^d	Production de niacine	Réduction de nitrate	
	TbD1+ RD4- RD7- RD8- RD9- RD10- RD12 ^{bov} RD13-	<i>M. bovis</i>	3, 16, 39-43	1	CTG (Leu)	ACC (Thr)	GAC (Asp)	Adénine	Δ6 pb	Sensible	Résistant	Négatif	Négatif	

Tableau 5. Caractéristiques génotypiques et biochimiques de *M. bovis*.

^a Lignées établies d'après Gagneux et Small (2007)

^b Lignées établies d'après SpolDB4 (Brudey et al., 2006)

^c TCH, hydrazide de l'acide thiophène carboxylique

^d PZA, Pyrazinamide

M. bovis présente un génome de 4,35 Mb, 65,6% de GC, 4 001 gènes et 33 pseudo-gènes. La présence de la région TbD1 et les allèles *KatG463* CTG (Leu) / *gyrA95* ACC (Thr), en fait une espèce ancestrale du groupe PGG1. En général, on ne retrouve qu'une seule copie d'IS6110 chez *M. bovis*, argument supplémentaire à son ancestralité. Le tableau 5 décrit les caractéristiques génotypiques et biochimiques de *M. bovis*.

2.3.4.3.4. Description de *Mycobacterium microti*

M. microti décrit il y a une soixantaine d'années (Wells et Oxon, 1937), fut d'abord nommée *Mycobacterium tuberculosis var. muris*. Cette première désignation est due à sa grande prévalence chez les muridées (famille appartenant à l'ordre des rongeurs ; Wells, 1953). Le nom *M. microti* fut attribué par Reed en 1957 (Reed, 1957). Même si cette espèce est très largement répandue chez les rongeurs, elle infecte également l'homme ; ce fait fut décrit pour la première fois par van Soolingen en 1998 (van Soolingen et al., 1998).

^a Lignées selon les LSP		^b Lignées selon le locus DR		IS6110 RFLP	Polymorphisme de séquence					Antibiotique		Caractéristiques biochimiques	
Nom	LSP	Nom	Espaceur absent	Nb bandes (Moy.)	<i>KatG463</i>	<i>gyrA95</i>	<i>pncA57</i>	<i>oxyR 285</i>	<i>pkc 15/1</i>	^c TCH (1µg/mL)	PZA (50 - 100 µg/mL) ^d	Production de niacine	Réduction de nitrate
	TbD1+ RD7- RD8- RD9- RD10- RD1 ^{mic}	<i>M. microti</i>	1-36, 39-43	13	CTG (Leu)	ACC (Thr)	CAC (His)	Guanine	Δ6 pb	Sensible	Sensible	Positif	Négatif

Tableau 6. Caractéristiques génotypiques et biochimiques de *M. microti*.

^a Lignées établies d'après Gagneux et Small (2007)

^b Lignées établies d'après SpolDB4 (Brudey et al., 2006)

^c TCH, hydrazide de l'acide thiophène carboxylique

^d PZA, Pyrazinamide

Au niveau du génome, les souches infectant les animaux et celles retrouvées chez les hommes ne présentent pas le même nombre de copies d'IS6110 (Cavanagh et al., 2002). Le profil des SNPs, *katG463*, *gyrA95* en fait une espèce appartenant au groupe PGG1, et présentant des caractères ancestraux, vu la présence de la région TbD1. Le tableau 6 décrit les caractéristiques génotypiques et biochimiques de *M. microti*.

2.3.4.3.5. Description de *Mycobacterium pinnipedii*

M. pinnipedii se retrouve en grande partie chez les pinnipèdes mais aussi chez d'autres animaux (exemple : le cochon, le lapin et le tapir). Il fut décrit pour la première fois par Cousins en 2003 (Cousins et al., 2003). *M. pinnipedii* est également transmissible à l'homme (Kiers et al., 2008).

^a Lignées selon les LSP		^b Lignées selon le locus DR		IS6110 RFLP	Polymorphisme de séquence					Antibiotique	Caractéristiques biochimiques		
Nom	LSP	Nom	Espaceur absent	Nb bandes (Moy.)	KatG463	gyrA95	pncA57	oxyR 285	pks 15/1	^c TCH (1µg/mL)	PZA (50 - 100 µg/mL) ^d	Production de niacine	Réduction de nitrate
TbD1+ RD7- RD8- RD9- RD10- RD2 ^{seal}		PIN1	1-3, 8- 22, 39- 43	5	CTG (Leu)	ACC (Thr)	CAC (His)	Guanine	Δ6 pb	Sensible +	Sensible	Négatif	Négatif
		PIN2	1-24, 39-43										

Tableau 7. Caractéristiques génotypiques et biochimiques de *M. pinnipedii*.

^a Lignées établies d'après Gagneux et Small (2007)

^b Lignées établies d'après SpolDB4 (Brudey et al., 2006)

^c TCH, hydrazide de l'acide thiophène carboxylique

^d PZA, Pyrazinamide

C'est une espèce ancestrale appartenant au groupe PGG1, comme les autres membres de MTC hors *M. tuberculosis* au sens strict. Le tableau 7 décrit les caractéristiques génotypiques et biochimiques de *M. pinnipedii*.

2.3.4.3.6. Description de *Mycobacterium mungi*

M. mungi fut récemment identifié chez les mangoustes du Botswana (Alexander et al., 2010) dont le lieu de vie est proche de celui des hommes. *M. mungi* partagerait un ancêtre avec MWA2. En effet, il ne possède pas les régions de différence TbD1, RD9, RD7, RD8, RD10. Il présente cependant une délétion qui lui est spécifique : RD1^{mon}. Récemment découvert, seule une étude apporte des informations à cette nouvelle espèce de MTC (Alexander et al., 2010).

2.3.4.3.7. Description de *Mycobacterium canettii*

M. canettii doit son nom à Georges Canetti, le microbiologiste qui l'a découverte en 1969. *M. canettii* fut caractérisé de façon plus détaillée bien plus tard, en 1997 par van Soolingen (van Soolingen et al., 1997) grâce à des souches en provenance de Somalie. D'après les divers cas rapportés cette espèce a une prévalence et une spécificité géographique pour les régions de la Corne de l'Afrique (Pfyffer et al., 1998 ; Miltgen et al., 2002). Cette espèce se caractérise par un faible nombre de bandes d'IS6110, Au niveau du locus DR, il est important de préciser que 26 espaceurs (du 69^{ème} au 94^{ème}) des 94 espaceurs chez MTC sont spécifiques à *Mycobacterium canettii* (van Embden et al., 2000). Cette espèce est aussi ancestrale et fait partie du groupe PGG1. Le tableau 8 décrit les caractéristiques génotypiques et biochimiques de *M. canettii*.

^a Lignées selon les LSP		^b Lignées selon le locus DR		IS6110 RFLP	Polymorphisme de séquence					Antibiotique		Caractéristiques biochimiques	
Nom	LSP	Nom	Espaceur absent	Nb bandes (Moy.)	KatG463	gyrA95	pncA57	oxyR 285	pks 15/1	^c TCH (1µg/mL)	PZA (50 - 100 µg/mL) ^d	Production de niacine	Réduction de nitrate
TbD1+, RD12 ^{can}		CANETTII	1-29, 31-35, 37-43	2	CTG (Leu)	ACC (Thr)	CAC (His)	Guanine	Type sauvage	Résistant	Sensible	Positif	Positif

Tableau 8. Caractéristiques génotypiques et biochimiques de *M. canettii*.

^a Lignées établies d'après Gagneux et Small (2007)

^b Lignées établies d'après SpolDB4 (Brudey et al., 2006)

^c TCH, hydrazide de l'acide thiophène carboxylique

^d PZA, Pyrazinamide

2.3.4.3.8. Description de *Mycobacterium tuberculosis*

M. tuberculosis est essentiellement pathogène pour l'Homme. Les caractéristiques génotypiques et biochimiques de *M. tuberculosis* sont décrites dans le tableau 9. Cette espèce est subdivisée en plusieurs groupes phylogénétiques grâce à la classification spoligotypique basée sur le locus DR

(Brudey et al., 2006) : East African-Indian (EAI), Haarlem, Latin American-Mediterranean (LAM), X, S et la famille T mal définie. Ceux-ci sont soit ancestraux, soit modernes, et appartiennent à l'un des trois groupes décrit par Sreevatsan et al. en 1997. Les souches de cette espèce provoquent quasiment toujours une tuberculose pulmonaire. L'émergence des différentes espèces au sein de MTC est la résultante d'un processus évolutif. Ainsi, il est judicieux à ce point de l'introduction de décrire l'évolution en biologie.

^a Lignées selon les LSP		^b Lignées selon le locus DR			IS6110 RFLP	Polymorphisme de séquence					Antibiotique		Caractéristiques biochimiques	
Nom	LSP	Nom	Espaceur absent	Nb bandes (Moy.)	KatG463	gyrA95	pncA57	oxyR 285	pks 15/1	^c TCH (1µg/mL)	PZA (50 - 100 µg/mL) ^d	Production de niacine	Réduction de nitrate	
Indo-Oceanic	TbD1+	EAI	29-32, 34	variable	CTG (Leu)	ACC (Thr)	CAC (His)	Guanine	Δ 7 pb	Résistant	Sensible	Positif	Positif	
East Asian	TbD1-RD207-RD105-	Beijing	1-34	17										
East-African Indian	TbD1-RD750	CAS	4-7, 23-34	15										
Euro American	TbD1-	Haarlem, LAM, X, S, T	33-36	variable	CGG (Arg) ou CGG (Arg)	ACC (Thr) ou AGC (Ser)								

Tableau 9. Caractéristiques génotypiques et biochimiques de *M. canettii*.

^a Lignées établies d'après Gagneux et Small (2007)

^b Lignées établies d'après SpolDB4 (Brudey et al., 2006)

^c TCH, hydrazide de l'acide thiophène carboxylique

^d PZA, Pyrazinamide

3. Notion d'évolution en biologie

3.1. Les théories de l'évolution des êtres vivants

3.1.1. Le transformisme de Lamarck ou l'influence du milieu dans la complexification des êtres vivants

Jean-Baptiste Lamarck (1744-1829) est le premier évolutionniste à avoir émis l'idée que l'environnement avait un rôle dans la spécialisation des espèces. En

effet, Lamarck pense que ce n'est point la forme du corps des animaux qui influe sur l'habitude mais que c'est plutôt l'habitude qui modèle le corps par transformations graduelles au cours du temps. Voilà l'idée qui émerge de cette citation tirée de son discours inaugural de 1800 au Muséum National d'Histoire Naturelle : « *Je pourrais prouver que ce n'est point la forme du corps, soit de ses parties, qui donne lieu aux habitudes, à la manière de vivre des animaux ; mais que ce sont au contraire les habitudes, la manière de vivre et toutes les circonstances influentes qui ont avec le temps constitué la forme des animaux* ». Ainsi, de façon progressive, les organismes se complexifient, ce qui selon lui serait un processus naturel. Sa théorie de l'évolution repose sur deux lois :

1. *"Dans tout animal qui n'a point dépassé le terme de ses développements, l'emploi plus fréquent et soutenu d'un organe quelconque fortifie peu à peu cet organe, le développe, l'agrandit et lui donne une puissance proportionnée à la durée de cet emploi, tandis que le défaut constant d'usage de tel organe l'affaiblit insensiblement, le détériore, diminue progressivement ses facultés et finit par le faire disparaître."*

2. *"Tout ce que la nature a fait acquérir ou perdre aux individus par l'influence constante des circonstances où leur race se trouve depuis longtemps exposée, et par conséquent par l'influence de l'emploi prédominant de tel organe, ou par celle d'un défaut d'usage constant de telle partie, elle le conserve pour la génération de nouveaux individus qui en proviennent, pourvu que les changements acquis soient communs aux deux sexes, ou à ceux qui ont produit ces nouveaux individus."*

Ces deux lois suggèrent que la fonction crée l'organe. La seconde loi de Lamarck introduit la notion d'hérédité dans les connaissances scientifiques. Cette dernière notion non relevée par la communauté scientifique de son époque, trouvera largement sa place sous les lumières de Gregor Mendel (1822-1884), botaniste autrichien, qui établira les lois de l'hérédité quelques décennies plus tard.

3.1.2. La sélection naturelle de Charles Darwin et de Alfred Russel Wallace motrice de l'évolution biologique

Alfred Russel Wallace (1823-1913) fut le premier à croire à la transmutation des espèces. En 1858, dans un essai nommé « Sur la tendance des variétés à se

démarquer indéfiniment du type original », Wallace décrit une théorie de l'évolution basée sur la sélection naturelle. A cette même date Charles Darwin (1809-1882), affinait ses travaux basés sur cette même théorie de l'évolution des espèces grâce à la sélection naturelle, travaux amorcés en 1838. C'est d'ailleurs Darwin qui emploie le terme sélection naturelle pour la première fois dans son œuvre « De l'origine des espèces » apparue en 1859 (Darwin, 1859). Dans ce livre, il explique la transmutation des espèces par la pression de sélection qui mène à la conservation (ou survie) des espèces les mieux adaptées à l'environnement. Ainsi pour lui, l'adaptation est le fruit du hasard, et les variations environnementales contribuent directement à la diversification des espèces. Son ouvrage « De l'origine des espèces » pose les fondements de l'évolution qui seront consolidés plus tard par la génétique. Même si les notions génétiques étaient absentes à l'époque, Darwin a tout de même suggéré la transmission des caractères acquis tout en ignorant le support.

3.1.3. La biogéographie discipline introduite par Alfred Russel Wallace

La biogéographie, discipline vaste, introduite par Alfred Russel Wallace, est l'étude de la distribution géographique des espèces. Ce domaine scientifique se trouve au carrefour de bien d'autres disciplines que la biologie telles la géologie, la climatologie, et la météorologie. En gros cela implique tous les domaines relatifs aux divers facteurs ayant d'une part amené un organisme dans un milieu donné et d'autre part à ceux ayant permis son adaptation. Par exemple, l'étude de la biogéographie des groupes phylogénétiques identifiés au sein de *M. tuberculosis*, espèce exclusivement pathogène à l'homme, au-delà de l'étude des différents facteurs impliqués dans les interactions hôte-pathogène, s'étend indubitablement à l'étude des migrations humaines ayant façonné les différentes populations du monde.

Comme le montre la figure ci-dessous, Hershberg et al., en 2008 ont pu superposer la distribution géographique des lignées MTC aux grands mouvements migratoires humains. L'histoire migratoire humaine qui ressort de la distribution des lignées de MTC est en adéquation avec l'hypothèse du « out of africa ». Cette étude fut réalisée à l'aide du séquençage de 89 gènes chez 108 souches distinctes, base de données

qui a d'ailleurs permis d'inférer un arbre phylogénétique traduisant les relations existant entre les 6 lignées majeures de MTC.

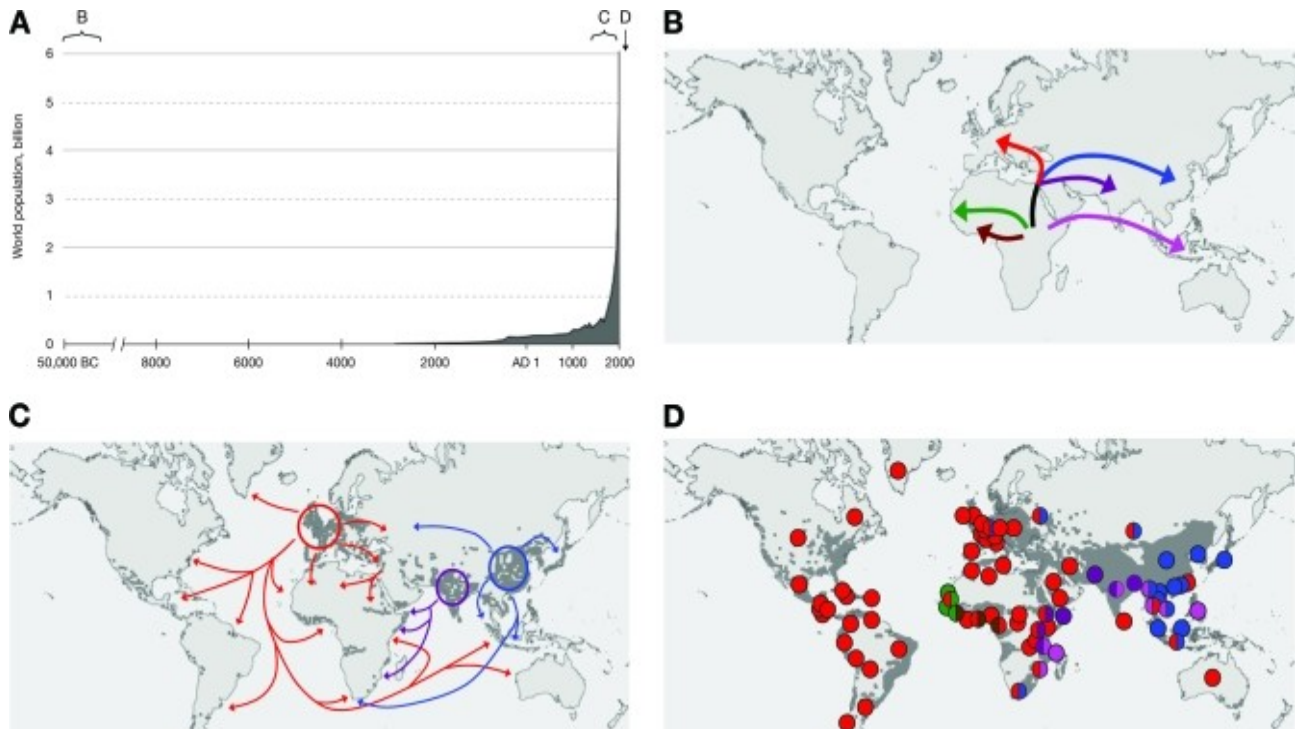


Figure 7. Image tirée de l'étude de Hershberg et al., en 2008. (A) Evolution de la taille de la population mondiale. (B) Migration hypothétique des lignées anciennes de MTC hors de l'Afrique (out of Africa). Indo-Oceanic (EAI) en violet clair, *M. africanum* West African 1 (AFRI2/3) en marron, *M. africanum* West African 2 (AFRI1) en vert, East Asian (Beijing) en bleu, East-African Indian (CAS) en violet foncé, Euro American (LAM, Haarlem, S, X, T) en rouge. L'ancêtre commun de East Asian, East-African Indian et Euro American en noir. (C) Augmentation récente de la population humaine globale (variation de gris). L'Inde, l'Est de l'Asie et l'Europe ont connu une forte augmentation de leur population et chacune de ces régions est associée à une des trois lignées phylogénétiques représentées. Des migrations récentes (commerce, colonisation) ont permis l'expansion de ces lignées modernes. (D) La population mondiale atteint 6 milliards d'habitants. La carte illustre la distribution mondiale actuelle des six lignées majeures de MTC pathogènes à l'homme.

3.2. Les méthodes de reconstruction phylogénétique pour l'inférence des relations de parenté inter/intra-espèces

Au cours de l'évolution, grâce aux mutations qui affectent le génome et à la sélection naturelle, chaque espèce appartenant à un même genre, s'individualise par une signature génomique qui lui est propre. Les marqueurs moléculaires qui ciblent des régions polymorphiques spécifiques, attestent des signatures génomiques pour pouvoir discriminer les espèces entre elles. Une fois les espèces identifiées, le recours à une méthode d'inférence phylogénétique est une étape indispensable à la

mise en évidence des liens phylogénétiques existants entre espèces. Les méthodes cladistiques établissent les relations de parenté en considérant les états de caractères partagés (homologie).

3.2.1. La parcimonie de Wagner

Les méthodes de parcimonie sont issues de la cladistique. Elle part du principe que l'évolution est parcimonieuse puisque qu'elle requiert un minimum de changements évolutifs. Sur un plan moléculaire, l'arbre optimal est celui qui minimise le nombre global de mutations, ainsi les branches les plus courtes sont privilégiées. De même, cet arbre minimise les phénomènes de conversion évolutive et de réversion. Cette méthode requiert 3 étapes : la recherche de tous les cladogrammes possibles, la mesure de la longueur de chaque arbre et enfin la sélection de ceux présentant la plus petite longueur. Les arbres obtenus peuvent être polarisés en considérant des espèces externes (outgroups).

3.2.2. La méthode bayésienne : une approche probabiliste

L'approche bayésienne en phylogénie, formalise le calcul des probabilités postérieures d'arbres phylogénétiques en fonction de probabilités connues à priori. La probabilité postérieure renvoie à la probabilité de l'hypothèse H sachant les données X. Cette probabilité postérieure d'une l'hypothèse est calculée grâce au théorème de Bayes : $P(H|X)=P(X|H).P(H)/P(X)$. Elle est donc fonction de la fonction de vraisemblance $P(X|H)$ (probabilité d'observer X sous l'hypothèse H), de $P(H)$ la probabilité à priori de H, et de $P(X)$ qui est la probabilité d'observer les données. Appliquer cette formule à la phylogénie revient donc à calculer la probabilité postérieure d'un arbre T_i sachant les données X suivant la formule :

$$f(\tau_i|X) = \frac{f(X|\tau_i)f(\tau_i)}{\sum_{j=1}^{B(s)} f(X|\tau_j)f(\tau_j)}$$

Au-delà des données, la probabilité postérieure est aussi fonction d'autres paramètres relatifs aux longueurs des branches et au modèle d'évolution puisque la fonction de vraisemblance se décrit comme suit :

$$f(X|\tau_i) = \int_v \int_{\theta} f(X|\tau_i, v, \theta) f(v, \theta) dv d\theta.$$

La sommation (relative au dénominateur de la précédente formule mathématique) concerne tous les arbres possibles pour s taxons ou $B(s)$. L'intégration est fonction de toutes les combinaisons de longueurs de branches v et des paramètres de substitution θ . La probabilité à priori d'un arbre $f(\tau_i) = 1/B(s)$. La fonction de vraisemblance $f(X|\tau_i, v, \theta)$ est calculée en admettant que les substitutions arrivent suivant un processus de Poisson homogène. Cependant, la sommation et les intégrations requises pour l'analyse bayésienne sont analytiquement impossibles. Finalement, on estime la probabilité postérieure par l'approche de Markov chain Monte Carlo (MCMC) (Metropolis and Ulam, 1949 ; Lewis, 2001). C'est d'ailleurs la méthodologie utilisée par le logiciel d'inférence phylogénétique MrBAYES (Huelsenbeck et Ronquist, 2001 ; Ronquist et Huelsenbeck, 2003). Le MCMC estime la distribution de probabilité en échantillonnant les valeurs des paramètres de façon périodique. On a l'habitude de considérer l'espace multidimensionnel des paramètres comme un terrain de randonnée sur lequel évolue de façon aléatoire un randonneur. Il couvrira la totalité de l'espace s'il marche assez longtemps, ce qui lui donnera une idée de la topographie du terrain. Cette représentation de la topographie sera d'autant plus précise si au préalable il se fixe des règles d'évolution dans sa marche : (1) s'il peut faire un pas en montant, il le fait toujours, (2) si un pas lui fait perdre de l'altitude, il ne le fait systématiquement. Cela le poussera à visiter fréquemment les points les plus élevés. En supprimant le début de sa trace (étape d'allumage), considéré comme l'étape d'approche, on pourra apprécier (avec une précision qui sera fonction du temps passé à marcher) la forme du paysage.

Revenons aux explications mathématiques. Un processus Markovien est un processus stochastique dans lequel la prédiction du futur dépend uniquement de l'état des variables au présent. Les faits antérieurs n'ont aucune incidence sur l'avenir. Soit, l'espace X où $x = \{x_0, x_1, \dots, x_n, x_{n+1}, \dots\}$ est une séquence d'états correspondant à la réalisation de d'une chaîne de Markov sur X . La probabilité $P(x_{n+1} | x_0, x_1, \dots, x_n)$ est égale à $P(x_{n+1} | x_n)$. Or la probabilité d'être dans l'état X_{n+1} en sortant X_n en faisant un pas est :

$$P_{n+1}(x_{n+1}) = \sum_{x_n} P_n(x_n)Q(x_n, x_{n+1})$$

Où Q est la matrice de transition, donc $Q(x_n, x_{n+1})$ est la probabilité de faire une transition de l'état x_n à x_{n+1} . $P_n(x_n)$ est la probabilité d'observer x_n .

La chaîne de Markov est reproduite à l'issue des nombreuses itérations de l'algorithme de Metropolis Hastings (Metropolis et al., 1953 ; Hastings, 1970). Soit $\psi = (\mathcal{T}, v, \theta)$, un arbre, un vecteur de longueurs de branches, les paramètres d'évolution. L'algorithme part d'un état initial ψ et propose un nouvel état ψ' sachant l'état initial. Le nouvel état est accepté avec une probabilité R. Si R est supérieure à une variable aléatoire uniforme entre 0 et 1, l'état ψ' est accepté ($\psi = \psi'$). A l'inverse ψ' est rejeté ($\psi = \psi$).

Une des limites des MCMC est le défaut de convergence de ces chaînes. Les chaînes se retrouvent dans ce cas piégées dans un sommet alors que les autres n'ont pas encore été visités. Pour éviter le piègeage des MCMC dans des optimaux locaux, le « Metropolis Coupling Markov Chain Monte Carlo » appelé simplement MC³ utilise plusieurs MCMC de façon simultanée. Parmi les n MCMC utilisées, n-1 dites chaudes explorent plus largement l'espace grâce à un pas plus large. Ces chaînes chaudes guident l'unique chaîne froide qui est utilisée pour les inférences. A chaque génération, les sauts entre les états des chaînes chaudes et ceux de la chaîne froide permettent à cette dernière de sortir d'un optimum local éventuel. Cette tournure algorithmique du MC³ diminue le risque de défaut de convergence des MCMC. La méthode d'utilisation du logiciel MrBayes sera décrite dans la partie méthodologie.

3.2.3. Les méthodes de distances

3.2.3.1. Unweight Pair Group Method with Arithmetic mean (UPGMA)

Cette méthode utilise une technique d'agglomération hiérarchique pour construire un arbre phylogénétique enraciné. Elle utilise obligatoirement une matrice de distances où tous les taxa sont comparés deux à deux. A chaque itération, les deux organismes présentant la plus faible distance sont placés dans l'arbre et portés par un même nœud. Dans la matrice, ils sont remplacés par un élément consensus, à partir duquel des nouvelles distances sont calculées. Les deux organismes les plus

proches vont ainsi être placés dans l'arbre, à chaque itération jusqu'à ce que la matrice soit vide.

L'inconvénient de la méthode, c'est qu'elle suppose que la vitesse d'évolution soit constante pour tous les organismes considérés. Si un organisme évolue plus vite, il va se retrouver à l'extérieur de l'arbre : phénomène de l'attraction des longues branches. La racine de l'arbre est placée à égale distance entre les deux organismes placés à la dernière itération.

3.2.3.2. Le Neighbor-Joining

La méthode du Neighbor-Joining est superposable aux autres méthodes d'analyse de groupe. Comme l'UPGMA, elle procède par itération mais aboutit à un arbre non enraciné. Dans ce cas, on part d'un nœud central, puis par itérations successives, et de proche en proche les organismes sont placés dans l'arbre. Dans cette méthode, on admet une vitesse d'évolution variable ce qui se traduit sur l'arbre par des branches de longueurs variables.

3.2.3.3. Le minimum spanning tree

Le Minimum Spanning Tree (MST) est une méthode de distances puisqu'il utilise une matrice de distance pour définir la topologie d'un arbre. Cependant, il obéit au principe du maximum de parcimonie (MP) en admettant que l'évolution est parcimonieuse. Ainsi, l'arbre qui résulte de cette technique présente une topologie où la somme des longueurs des branches est minimale. Si le MST et le MP se partagent le même principe de base, ils présentent bien évidemment des différences qui font d'eux des méthodes distinctes qui n'aboutissent pas à la même topologie pour un même échantillon donné. Contrairement au MP, dans la définition du MST, il est exclu d'admettre des unités ancestrales (correspondant aux nœuds internes de l'arbre) qui ne fassent pas partie de l'échantillon : on parle alors de nœuds hypothétiques. Le nœud hypothétique est un ancêtre que l'on suppose avoir existé mais qui est absent de l'échantillon étudié. Son existence est supposée du fait que les critères de la parcimonie requièrent sa présence comme un intermédiaire obligatoire pour ajouter de la cohérence aux états recueillis et de ce fait arriver à une topologie qui explique la réalité observée. Pour pallier à ce refus d'admettre des nœuds hypothétiques, le MST réclame un jeu de données exhaustif reflétant donc

une biodiversité sans lacune. Pour satisfaire cette exigence, l'étude doit s'étendre sur une courte durée, pour laquelle l'ensemble des états (relatifs à la région polymorphique ciblée par le marqueur considéré) soit encore vérifiable. Un échantillon incomplet amène à l'absence de pas évolutif dans l'histoire évolutive retracée. Si trop d'états sont absents de l'échantillon, l'histoire évolutive qui en ressortira après soumission à un algorithme MST ne sera pas du tout superposable à l'histoire réelle et ce sur toute sa longueur. Une autre différence du MP c'est qu'il conçoit qu'un nœud ne puisse porter que deux branches ; un ancêtre possède uniquement deux descendants. Le MP génère donc des arbres dichotomiques. Ce n'est pas le cas du MST. Le nœud de départ (nœud racine) choisi par le MST est l'état qui présente le plus d'états liés ou de descendants directs et à partir de celui-ci il dérive les autres états. Il en résulte un arbre avec des branches en étoile. Cette méthode fut très utilisée au cours de cette thèse grâce au logiciel BioNumerics, dont l'algorithme MST qu'il propose sera décrit dans la partie méthodologie.

4. Evolution phylogénétique de MTC

4.1. Histoire évolutive du complexe *Mycobacterium tuberculosis*

Une histoire évolutive illustrant les relations phylogénétiques entre les six lignées majeures de MTC est présentée par l'arbre phylogénétique qui suit. Cet arbre phylogénétique résulte de la méthode du Maximum de Parcimonie de 108 souches *M. tuberculosis* en utilisant 89 séquences de gènes distincts (Hershberg et al., 2008). L'arbre de Hershberg fut agrémenté par de Jong et al., en 2010 en ajoutant les informations relatives aux régions de délétion (RD) et aux SNPs (figure 8).

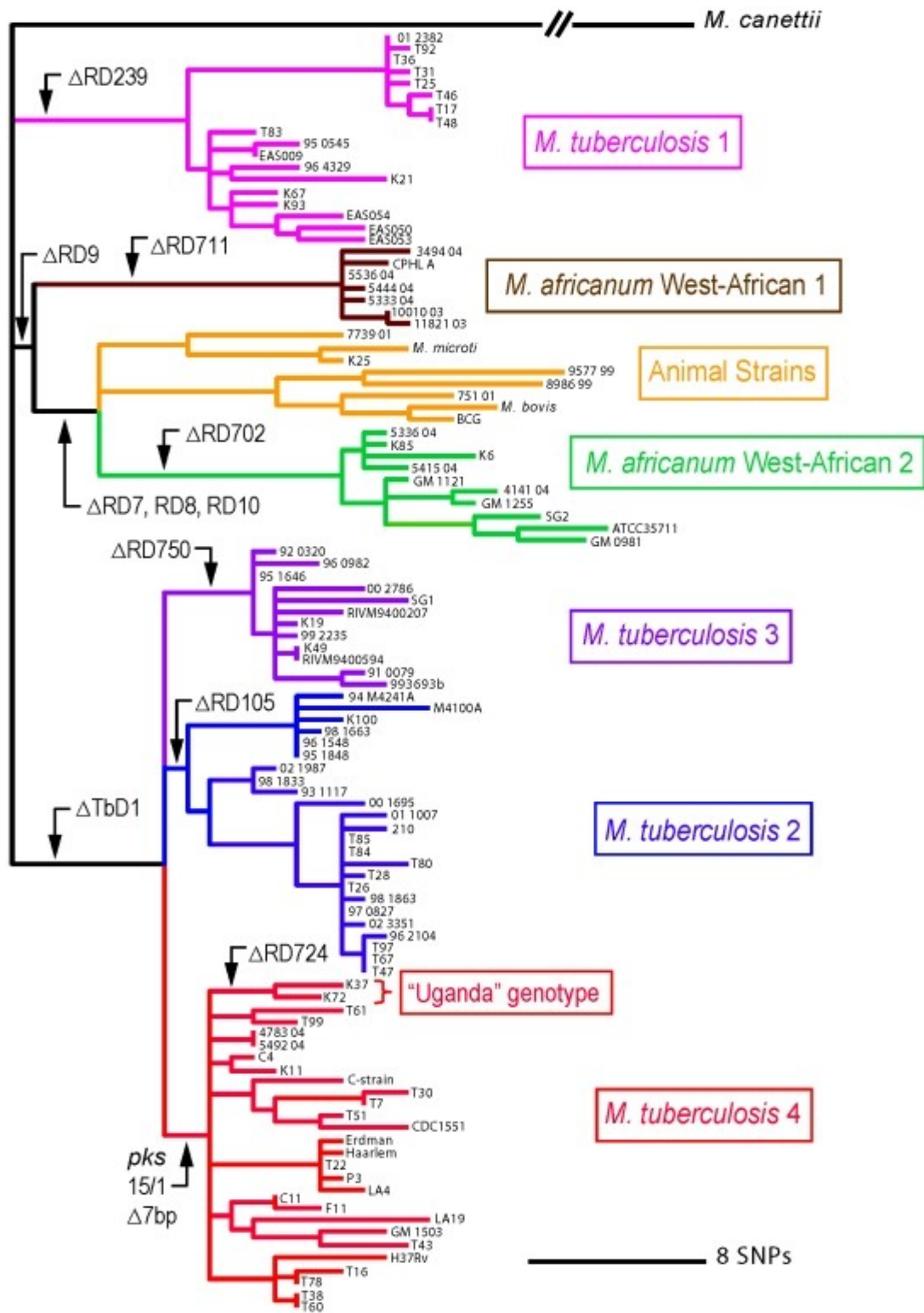


Figure 8. Arbre phylogénétique résultant de la méthode du Maximum de Parcimonie de 108 souches *M. tuberculosis*. Les flèches noires marquent les points de survenance des régions de différence (RD). L'échelle est établie en fonction du nombre de SNPs. *M. tuberculosis* 1 (rose) correspond à Indo Oceanic (EAI), *M. tuberculosis* 2 (bleu) correspond à East-African Indian (CAS), *M. tuberculosis* 3 (violet) correspond à East Asian (Beijing) et *M. tuberculosis* 4 (rouge) correspond à Euro American (LAM, Haarlem, X, S et T).

D'après Gutierrez et al., (2005), *M. canettii* serait la plus ancienne espèce de MTC. Ils ont démontré que *M. canettii* fait partie du groupe des mycobactéries lisses duquel tous les autres membres de MTC, auraient émergé. Ceci explique le fait que *M. canettii* soit considéré comme un outgroup proche des autres membres de MTC. *M. tuberculosis* 1 (EAI) serait le premier à diverger suivi de l'ancêtre d'un groupe formé par *M. africanum* (West-African 1 et 2) et des lignées animales (tels *M. bovis*, *M. microti*). *M. Africanum* West-African 2 (MAW2) est très proche des lignées animales, ce qui laisse supposer un réservoir animal pour les souches de MAW2 (de Jong et al., 2010). Bien après, l'ancêtre de l'ensemble des lignées appartenant à *M. tuberculosis*, à l'exception de EAI/Indo-Oceanic, aurait divergé. Cet ancêtre aurait ensuite donné naissance à 3 groupes phylogénétiques distincts : East-African Indian/CAS, East Asian/Beijing, et Euro American/LAM, Haarlem, X, S, T. Ces trois groupes phylogénétiques composent les lignées modernes par opposition aux lignées ancestrales qui regroupent les autres lignées phylogénétiques de MTC. Cet arbre issu d'un algorithme MP à partir de 89 séquences distinctes chez 108 souches fut conforté par les régions de différences et les SNPs. Ceci montre à quel point la confrontation des marqueurs moléculaires est importante dans une étude donnée pour attribuer un poids aux hypothèses émises. Ainsi, l'exploration de la diversité génotypique des bacilles tuberculeux, par la mise en place de base de données où les profils d'origine géographique diverse sont décrits selon plusieurs marqueurs moléculaires, est essentielle pour asseoir les études phylogénétiques (mais aussi épidémiologiques) pour comprendre le MTC dans sa globalité. Il faut aussi cerner les potentialités des marqueurs, pour une utilisation juste et optimale de ces derniers.

4.2. Apport de la bioinformatique dans la reconstruction phylogénétique de MTC

4.2.1. La base de données de génotypes mycobactériens, outils indispensable pour comprendre l'évolution et la biogéographie de MTC

SpolDB4 est une base de données spoligotypiques (Brudey et al., 2006), contenant 39 295 souches provenant de 122 pays. Elle compte 1939 types partagés de spoligotypage (ou SIT) au format 43 espaceurs.

Un numéro de SIT est attribué aux patterns de spoligotype⁴³ présents au moins 2 fois dans la base de données. Cette accumulation de données a permis d'identifier des clusters de souches ayant en commun une signature spoligotypique, encore appelée règle de classification ou prototype. Ainsi, grâce à un expert et une approche informatique, 62 clades (lignées) et sous-clades furent identifiés, comme le montre la figure 9.

Comme l'a montré Hershberg et al., en 2008, les lignées de MTC ont une spécificité géographique. Par exemple, Central-Asian (CAS) comprend deux sous-lignées CAS1-Delhi que l'on observe majoritairement en Inde (Bhanu et al., 2002) et CAS1-Kilimanjaro prévalent en Tanzanie (McHugh et al., 2005). La famille East-African-Indian (EAI) a une géospécificité pour la côte africaine et le Sud de l'Asie. Cependant, les sous familles de EAI présentent une géospécificité plus réduite à l'intérieure de ces zones cibles. EAI1-SOM se retrouve sur la côte Est africaine, EAI3-IND est prévalent en Inde.

4.3. Outils bioinformatiques pour l'acquisition de nouvelles connaissances sur le complexe *Mycobacterium tuberculosis*.

4.3.1. Les réseaux bayésiens, une technique de l'intelligence artificielle pour extraire les informations cachées dans les grandes bases de données

Les réseaux bayésiens sont les résultats du travail de Thomas Bayes, sur la probabilité des causes. Ceux-ci sont des modèles graphiques unissant la théorie des probabilités à celles des graphes afin de structurer simplement un système complexe et d'établir une cohérence des données incertaines au sein de ce système. C'est une technique très intéressante pour représenter clairement une quantité importante de données. Il convient pour cela de définir les variables d'intérêt qui décrivent le système. Ensuite d'établir les relations d'influence qui existent entre les variables. Ainsi, le graphe obtenu (ou structure du modèle) se compose de nœuds (symbolisant les variables) et de flèches (pour les relations de causalité). À chaque nœud (ou variable) sont associées des tables de probabilité. L'utilisation d'un réseau bayésien repose donc sur la propagation de l'information dans le modèle construit,

autrement dit sur le calcul de la probabilité d'un évènement connaissant certaines observations. C'est ce que l'on appelle l'inférence. L'inférence permet le traitement de requêtes complexes. On peut vérifier la valeur la plus probable d'un nœud, la cause la plus probable d'une information, on peut évaluer la sensibilité à une information en faisant varier l'état de la variable en question et en observant parallèlement les variations de la probabilité d'une hypothèse donnée. Il devient ainsi possible d'acquérir des connaissances nouvelles sur un système, connaissances contenues dans les données mais qui restent imperceptibles sans l'intervention de techniques de l'intelligence artificielle.

4.3.2. Importance d'un modèle bayésien quant à la classification des souches du complexe *Mycobacterium tuberculosis*

Le typage basé sur les minisatellites MIRU-VNTR a d'abord été appliqué sur un format de 12 loci (Supply et al., 2006) parmi les 41 loci trouvés dans MTBC génome, puis ensuite les formats 15 et 24 (Supply et al., 2000 ; Supply et al., 2001). Les minisatellites MIRU-VNTR ont un pouvoir discriminant supérieur au spoligotypage. Ces 2 marqueurs sont souvent utilisés en combinaison pour éviter des erreurs de classification en raison du phénomène de convergence évolutive (Warren et al., 2002 ; Abadia et al., 2010). Ces deux marqueurs qui n'ont ni la même horloge moléculaire (Supply et al., 2000) ni le même processus évolutif (Warren et al., 2002 ; Thorne et al., 2007 ; Billamas et al., 2007) constituent les piliers de « SITVIT2 » (qui est la version en cours de la base de données génotypiques). Le locus DR (Direct Repeat), région génomiques cible du spoligotypage présente un taux de mutation plus bas que les loci VNTR. De ce fait, le spoligotypage fut jugé plus approprié pour la classification des souches MTC. Alors que les minisatellites VNTR sont plutôt utilisés pour mettre en évidence les transmissions récentes, ou pour générer des grappes plus discriminantes dans des études spécifiques. Par exemple des études avec une forte prévalence de souches de la famille Beijing. En effet, le profil spoligotypique de Beijing a peu de marge de variabilité donc des souches peu discriminées entre elles. L'intérêt porté à ces deux types de marqueurs est différent bien qu'ils soient souvent utilisés seuls ou de façon combinée pour reconstruire l'histoire évolutive de MTC. Cependant, une inférence phylogénétique optimale dépendra : (i) du choix d'une méthode appropriée, (ii) du type de données

utilisées (séquences d'acides nucléiques, données catégoriques ou binaire), (iii) du taux d'évolution des différents taxons, (iv) de l'étalement de la période de l'étude, et (v) de l'exhaustivité de l'échantillon d'étude (qui elle-même dépend de la période évolutive ciblée). Afin d'avoir une interprétation juste de l'histoire évolutive proposée, il est aussi très important de comprendre les mécanismes mutationnels conduisant au polymorphisme de la région génomique ciblée par les marqueurs moléculaires. Malheureusement, les mécanismes à l'origine du polymorphisme du nombre de répétitions des unités minisatellites ainsi que du polymorphisme du locus DR sont encore inconnus. Par exemple, même s'il est largement admis que le locus DR évolue par la perte d'espaceurs, il reste difficile de savoir si un événement mutationnel implique la perte d'un espaceur unique et / ou de blocs d'espaceurs.

La base de données « SITVIT2 » constitue un grand échantillon de données sous-exploitées qui à l'aide de techniques de l'intelligence artificielle, peut apporter des informations nouvelles sur *M. tuberculosis*. L'intérêt d'un modèle bayésien est de structurer de façon cohérente les données afin de capturer les connaissances cachées dans cette énorme base de données internationales. A la lumière de ces nouvelles informations, il sera possible de soulever les cohérences et incohérences du schéma de classification spoligotypique et d'évaluer la capacité des loci VNTR à classifier les souches de MTC. Si cette classification est envisageable, il sera possible d'entrevoir la méthodologie à adopter pour la définir. Les réseaux bayésiens sont une alternative intéressante pour puiser les informations cachées dans les données de génotypage.

5. Les problématiques auxquelles doit répondre cette thèse

Cette thèse doit répondre à plusieurs problématiques :

- Décrire le paysage génotypique de MTC dans certaines régions de l'Afrique subsaharienne, de l'Inde et de la Chine, qui sont les régions les plus touchées par la tuberculose en 2009,
- Apporter à la base de données SITVIT2, des informations de typage au format MIRU-VNTR pour ces mêmes régions. En effet, la quantité de données collectées, n'est pas homogène partout dans le monde. Ce fait est bien plus marqué pour les informations de typage relatives à MIRU12 par rapport au spoligotypage. Par exemple, dans SITVIT2, l'Afrique de l'Est ne présente pas

d'isolats MIRUtypés,

- Etablir un schéma de classification basé sur les minisatellites MIRU-VNTR (format 12 loci),
- Confronter cette nouvelle classification à celles déjà établies et basées sur les LSP et le spoligotypage,
- Illustrer la distribution des lignées MIRU-VNTR dans les différentes sous-régions du monde,
- Développer un modèle bayésien qui décrit MTC, et qui permet (i) l'attribution des lignées à partir des spoligotypes et/ou des patterns MIRU12, (ii) la complétion des informations de typage, et (iii) l'attribution d'une origine géographique,
- Faire une analyse bioinformatique des séquences MIRU-VNTR.

Les résultats apportés visent à enrichir la diversité génotypique de MTC dans la base de données SITVIT2, à connaître mieux les marqueurs moléculaires utilisés et à comprendre mieux l'évolution de *Mycobacterium tuberculosis*, depuis son apparition sur le continent africain, il y a plusieurs milliers d'années.

MATÉRIELS ET MÉTHODES

1. Les méthodes de génotypage moléculaire utiles à la description des souches MTC

Le génome des mycobactéries, comme tout autre génome, connaît des variations génétiques qui peuvent être maintenues par la descendance, si elles n'entravent pas l'intégrité phénotypique qui assure la persistance de la bactérie dans son milieu. Certaines de ces régions polymorphiques sont ciblées par des marqueurs moléculaires dans le but de suivre la transmission de ces mutations génétiques d'une génération à une autre. L'utilisation des marqueurs moléculaires est indispensable pour la caractérisation des organismes au cours des études épidémiologiques et phylogénétiques. Ainsi, plusieurs méthodes de génotypage ; IS6110-RFLP, spoligotypage, typage des minisatellites VNTR, sont actuellement utilisées pour (i) mettre en évidence la diversité génétique des souches appartenant au complexe *Mycobacterium tuberculosis*, (ii) comprendre les évènements épidémiologiques, (iii) et étudier l'histoire évolutive des différentes espèces de MTC.

1.1. Description de l'IS6110-RFLP (polymorphisme de longueur des fragments de restriction)

La séquence d'insertion IS6110 est décrite dans la partie 2.3.3.2.2 du premier chapitre (Introduction). Le marqueur IS6110-RFLP est fréquemment utilisé pour la discrimination des souches. Cette méthode permet la formation de clusters de souches ayant en commun un nombre de copies d'IS6110, Un profil identique d'IS6110, chez deux patients en contact, suggère une transmission récente. Un même profil d'IS6110, dans aux moins deux prélèvements issus d'épisodes tuberculeux distincts pour un même patient, traduit une réactivation. En dessous de 6 copies d'IS6110, la discrimination des souches doit obligatoirement passer par une autre méthode de génotypage. Le typage des minisatellites MIRU-VNTR est souvent utilisé en recours pour discriminer les souches présentant un faible nombre de copies d'IS6110 ($n < 6$ copies). En général, le nombre de copies d'IS6110 est très

élevé chez les souches *M. tuberculosis* (jusqu'à 25 copies d'IS6110), ce qui confère à cette technique un très fort pouvoir discriminant (Devallois et al., 1998) et lui a valu le titre de méthode de génotypage de référence pour les souches MTC. On distingue tout de même des isolats à faible nombre de copies (Goyal et al., 1999) ou sans IS6110 dans leur génome (Yuen et al., 1993). La plupart des isolats de *M. bovis* présentent un faible nombre de copies (en moyenne une copie d'IS6110) (Aranaz et al., 1996 ; Aranaz et al., 1998).

1.1.1. Principe général du polymorphisme de longueur des fragments de restriction

Après extraction, l'ADN est coupé à différents points spécifiques ou sites de restriction par une enzyme de restriction. Toute variation au niveau d'un site de restriction empêche l'action de l'enzyme. Après la digestion enzymatique, les fragments de restriction sont séparés par électrophorèse, puis visualisés grâce à une sonde fluorescente ou radioactive. L'intérêt de la méthode réside dans le polymorphisme de longueur et dans le nombre des fragments de restriction.

1.1.2. Description du typage par IS6110-RFLP

Le protocole de l'IS6110-RFLP (figure 1) utilise généralement la technique de *Southern Blot*. Il a été développé par (van Embden et al., 1993) et se réalise comme suit :

- L'extraction et la purification d'ADN d'une bonne quantité de souches de MTC,
- La digestion des acides nucléiques par l'endonucléase *PvuII*,
- La séparation des fragments de restriction sur gel d'agarose,
- Le transfert des fragments d'ADN sur une membrane de nylon,
- L'hybridation des fragments avec une sonde (marquée à la peroxydase) spécifique de l'IS6110,
- La révélation après exposition d'un film auto-radiographique.

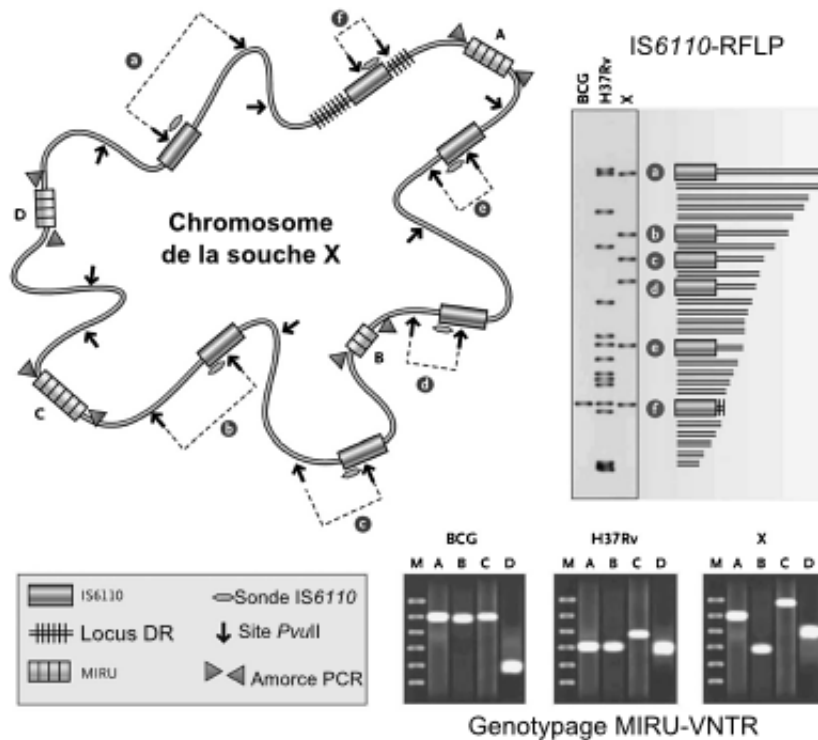


Figure 1. Génotypage d'une souche BCG, H37Rv et X (désignant une souche quelconque) avec les techniques IS6110-RFLP et MIRUs. En haut à gauche : génome d'une souche quelconque X ; en haut à droite : résultats d'IS6110-RFLP. L'ADN est digéré par l'enzyme de restriction *PvuII*. Les fragments obtenus sont visualisés sur gel d'agarose. En bas à droite : résultats de typage MIRU-VNTR : M = marqueurs de poids moléculaire ; A, B, C et D = résultats pour 4 loci MIRU (Tibayrenc, 2007).

1.2. Description des techniques ayant recours à la PCR

1.2.1. Le spoligotypage (Spacer Oligonucleotide Typing)

Le locus CRISPR est décrit dans la partie 2.3.3.2.4 du premier chapitre (Introduction). Il est encore appelé locus DR (Direct Repeat) et se compose de 94 espaceurs différents de 35 à 41 pb, séparés par une séquence répétée de 36 pb (van Embden et al., 2000). Ce locus connaît des délétions qui sont à l'origine de son polymorphisme chez MTC. Ainsi, l'état de l'ensemble des espaceurs présents dans le locus DR permet la discrimination des souches MTC.

1.2.1.1. Illustration du principe

Le spoligotypage consiste en la révélation de 43 espaceurs du locus DR. Les 43 espaceurs considérés sont les espaceurs occupant les positions génomiques 2 à 4, 12 à 15, 16 à 31, 34 à 42, 51 à 53, 62 à 65. Ces espaceurs considérés dans le spoligotypage43 sont numérotés de 1 à 43 (figure 2). Ils sont amplifiés par PCR grâce à des amorces complémentaires aux DR (ou Direct Repeat) (figure 2). Le

tableau 1 décrit la PCR menant à l'amplification des espaceurs. L'amorce Dra est biotinylée (5' biotine GGTTTTGGGTCTGACGAC 3') afin de marquer les amplicons et opère de façon conjuguée avec la Drb (5' CCGAGAGGGGACGGAAAC 3'). Les produits PCR sont ensuite soumis au protocole du spoligotypage.

A

Réactif	Volume pour 1 tube (µl)
Eau Ultra Pure	28
Tampon Mix (X)	10
MgCl ₂ (50mM)	7
Primers Dra (20pmol/µl)	1
Primers Drb (20pmol/µl)	1
dNTPs (25mM)	0.5
rTAQ polymérase (5U/µl)	0,125
ADN	2

B

94°C	10 min	} 35 cycles
94°C	30 secs	
60°C	1 min	
72°C	2 min	
72°C	10 min	

Tableau 1. A) Composition du Mix de la PCR du spoligotypage. **B)** Description des cycles de la PCR du spoligotypage

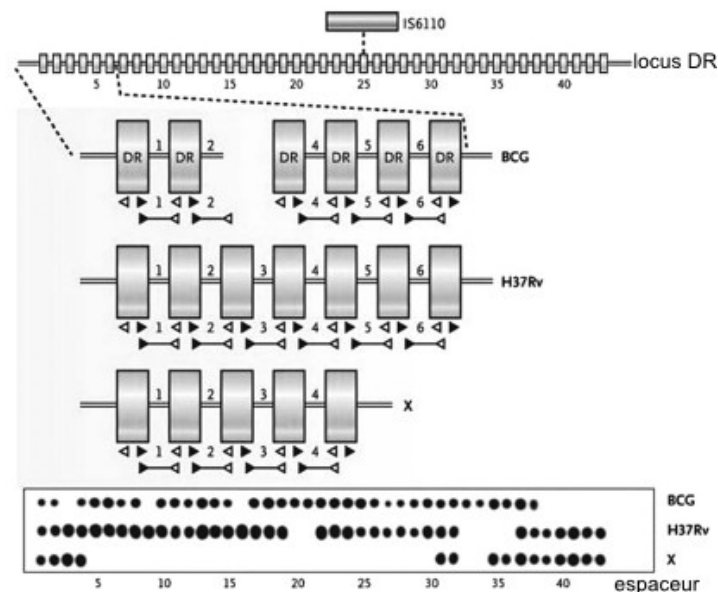


Figure 2. Le principe du spoligotypage. Tout en haut, une représentation du locus DR où les DRs (direct Repeats) sont intercalés par des espaceurs. Un IS6110 se situe dans un DR. La section du milieu montre la position des amorces Dra (biotinylé), DRb pour l'amplification des espaceurs présents chez les souches BCG, H37Rv et X. La dernière section présente les résultats du spoligotypage de BCG, H37Rv et X (Tibayrenc, 2007).

1.2.1.2. Description du protocole d'hybridation des produits PCR sur la membrane de nylon

Le protocole d'hybridation des produits PCR sur les sondes (spécifiques aux espaceurs) suit les étapes suivantes :

- La déshybridation par la chaleur des brins complémentaires des produits PCR,
- L'hybridation des échantillons d'ADN aux sondes fixées sur la membrane dans un miniblottter,
- Le lavage de la membrane pour enlever l'ADN non fixé à la membrane,
- La fixation de la streptavidine-péroxydase à la biotine portée par une des amorces des produits PCR,
- Le lavage de la membrane pour éliminer l'excédent de streptavine-péroxydase,
- La mise en contact de la membrane avec du luminol,
- L'exposition de la membrane contre un film photographique dans une cassette. Lors de la dégradation du luminol par la peroxydase, il y a libération d'énergie lumineuse (photons) qui imprime le film photographique hypersensible. Les sondes hybridées aux fragments amplifiés apparaissent alors sous la forme de points sombres sur le film (=présence d'espaceur), lors de l'étape de la révélation,
- Révélation du film photographique.

Les empreintes peuvent ensuite être traduites dans le format binaire ou octal.

1.2.1.3. Traduction des résultats dans des formats adéquats pour leurs manipulations

Le pattern original lu sur le film photographique est d'abord converti en binaire. Les points noirs attestent d'une présence d'espaceur et sont symbolisés par un 1 alors que les absences sont représentées par 0, Le format binaire se compose d'une série de 1 et de 0, Pour le passage du binaire à l'octal, les 43 caractères sont divisés en 14 groupes de 3 caractères et un groupe de 1 caractère (figure 3). On applique ensuite aux groupes de 3 caractères le codage suivant : 000 = 0; 001 = 1; 010 = 2; 011 = 3; 100 = 4; 101 = 5; 110 = 6; 111 = 7. Le dernier groupe qui se compose d'un seul caractère, garde sa valeur 0 ou 1.

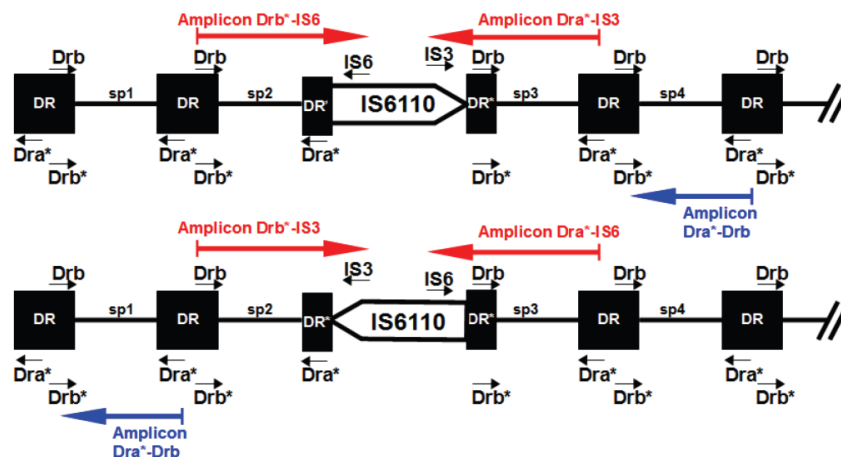


Figure 4. Illustration de la technique de génotypage pour la détection d'IS6110, Les amorces biotinylées sont suivies de « * ». Les espaceurs sont nommés « sp ». Dans la représentation du haut, l'IS6110 est insérée dans la direction 5'-3', alors que dans celle du bas, elle est dans la direction 3'-5'.

Cette technique fut appliquée sur toute la région portant les 86 premiers espaceurs du locus DR. Une empreinte sur le film photographique, révèle la présence d'un IS6110 dans l'un des deux DRs adjacents à l'espaceur révélé.

Deux membranes de nylon sont requises pour la manipulation. Chacune porte 43 sondes spécifiques à l'un des 86 espaceurs ciblés. La séquence nucléotidique de ces sondes est décrite par van Embden et al. (2000).

Comment interpréter les résultats ? Comme le montre la figure 4, une amplification des paires d'amorces (biot)DRa-IS3 (biot pour biotinylée), met en évidence une insertion d'IS6110 (dans la direction 5'-3') dans le DR qui précède l'espaceur amplifié. Une amplification des paires d'amorces (biot)DRb-IS6, met en évidence une insertion d'IS6110 (dans la direction 5'-3') dans le DR qui suit l'espaceur amplifié. Une amplification des paires d'amorces (biot)DRa-IS6, met en évidence une insertion d'IS6110 (dans la direction 3'-5') dans le DR qui précède l'espaceur amplifié. Une amplification des paires d'amorces (biot)DRa-IS6, met en évidence une insertion d'IS6110 (dans la direction 3'-5') dans le DR qui précède l'espaceur amplifié. Une amplification des paires d'amorces (biot)DRb-IS3, met en évidence une insertion d'IS6110 (dans la direction 3'-5') dans le DR qui suit l'espaceur amplifié.

Comme le montre la figure 4, les paires d'amorces (biot)DRa-IS3 et (biot)DRb-IS6 mettent en évidence la même IS6110, cela laisse penser qu'un seul des deux couples est suffisant. Il en est de même pour les couples (biot)DRa-IS6 et (biot)DRb-

IS3. Cependant, sachant que l'insertion de l'IS6110 peut être asymétrique dans le DR, cela peut compromettre la fixation de l'une des deux amorces complémentaires à ce DR, et compromettre également en aval, l'amplification d'un des deux espaceurs adjacents. Ainsi, il est nécessaire d'avoir recours aux deux couples d'amorces révélant la même insertion d'IS6110, pour réduire le risque de ne pas détecter l'insertion de la séquence d'insertion.

1.2.3. Les techniques basées sur l'amplification des minisatellites VNTR

Les VNTR sont des minisatellites répétés en tandem dispersés dans le génome des mycobactéries (Frothingham et al., 1998 ; Smittipat et al., 2000 ; Supply et al., 2000) et positionnés souvent entre deux gènes. Ils présentent un polymorphisme dans le nombre de répétitions. Chaque unité répétée compte entre 40 et 100 pb. La technique de génotypage basée sur les minisatellites VNTR est appliquée à MTC pour la discrimination des souches bactérienne d'une même espèce et étudier les routes de transmissions.

1.2.3.1. Description de la méthodologie

Le principe du typage par MIRU-VNTR est simple. Ce typage passe par les étapes suivantes :

- Préparation des Mix de la PCR. Dans chaque Mix est mis un couple d'amorces différentes. Les séquences des amorces sont disponibles en ligne (http://minisatellites.u-psud.fr/ASPSamp/Base_ms/VNTRprimers.php). La composition du Mix est décrite dans le tableau 2A.
- L'amplification des loci ciblés dans un thermocycleur, (voir les détails des cycles de la PCR dans le tableau 2B).
- La séparation des produits PCR par électrophorèse. La migration se fait durant 3 heures à 150 Volts, sur un gel d'agarose à 1,5%, dans une cuve remplie de TBE 1X (Tris-Borate EDTA).
- La révélation du gel après un bain de BET (0,5 mg/l) pendant 45 minutes environ.

A

Réactif	Concentration (solution mère)	Volume pour 1 tube (µl)	Concentration finale (MIX)
Eau Ultra-Pure		15,875	
Tampon Mix (X)	10	3	1
MgCl ₂ (mM)	25	2,4	2
Primers F (µM)	4	2	0,3
Primers R (µM)	4	2	0,3
DMSO (%)	100	3	10
dNTPs (mM)	25	0,6	0,5
rTAQ U/µl	5	0,125	0,02
TOTAL Mix (sans ADN)		29	
ADN	(0,02 µg/µl)	1	0,6 ng/µl

B

94°C	10 min	} 35 cycles
94°C	30 secs	
60°C	1 min	
72°C	2 min	
72°C	10 min	

Tableau 2. **A)** Composition du Mix de la PCR du typage MIRU-VNTR. **B)** Description des cycles de la PCR du typage MIRU-VNTR.

Comme le montre les trois gels de la figure 1, la position d'un fragment dans le gel révèle sa taille qui sera ensuite traduite en nombre de répétitions. L'ensemble de loci considérés, selon le format du marqueur MIRU-VNTR pour lequel on a opté, construit une empreinte génétique qui caractérise une souche.

1.2.3.2. Description des différents formats de loci considérés

Le premier format de typage MIRU-VNTR utilisé, rassemble 12 loci VNTR couramment appelés MIRUs (pour Mycobacterial Interspersed Repetitive Units) (Supply et al., 2001). Les formats 15 loci et 24 loci ont ensuite été proposés (Supply et al., 2006) (tableau 3).

Loci	Alias 1	Alias 2	24	15	12	Loci	Alias 1	Alias 2	24	15	12
154	MIRU02		X		X	2401	Mtub30		X	X	
424	Mtub04		X	X		2461	ETRB		X		
577	ETRC		X	X		2531	MIRU23		X		X
580	MIRU04	ETRD	X	X	X	2687	MIRU24		X		X
802	MIRU40		X	X	X	2996	MIRU26		X	X	X
960	MIRU10		X	X	X	3007	MIRU27	QUB5	X		X
1644	MIRU16		X	X	X	3171	Mtub34		X		
1955	Mtub21		X	X		3192	MIRU31	ETRE	X	X	X
2059	MIRU20		X		X	3690	Mtub39		X	X	
2163b	QUB11b		X	X		4052	QUB26		X	X	
2165	ETRA		X	X		4156	QUB4156		X	X	
2347	Mtub29		X			4348	MIRU39		X		X

Tableau 3. Description des formats 24 loci, 15 loci et 12 loci, utilisés lors du typage MIRU-VNTR. Tableau tiré du site MIRU-VNTRplus (www.miru-vntrplus.org).

Citons aussi, le marqueur VNTR qui se compose de ETRA, ETRB, ETRC, ETRD et ETRE.

1.2.4. Mise en place d'une nouvelle technique pour la détection des délétions adjacentes au point d'insertion d'IS6110

1.2.4.1. Principe

Pour déterminer si l'IS6110 est impliquée dans des recombinaisons génétiques qui entraîneraient des délétions adjacentes, phénomènes déjà décrits pour la séquence d'insertion IS10 (Roberts et al. 1991), la méthode de typage « IS6110 Adjacent Deletion » (IS6110AD-typing) a été développée dans le cadre de cette thèse. Cette technique vise à amplifier, par PCR, une partie des régions génomiques s'étalant sur 1500 pb autour du point d'insertion des IS6110, pour vérifier s'il y a une variabilité génomique à ces endroits.

ID IS6110	Position IS6110 (pb)	Nom amplicon	Amorce 1 (A1) 5'-3'	Position A1 (bp)	Tm A1	Amorce 2 (A2) 5'-3'	Position A2 (bp)	Tm A2	Taille amplicon (bp)	
Multiplex 1	IR-r start (pb)	889022	1-r	CGCCGACGATCGCAA	888041	62.71	TGAGGTTGGGCCGGT	888997	62.86	956
	IR-l end (pb)	890374	1-l	CCAACCTCGGTGAGAAAGACC	890455	62.24	CCTTTGGTAGTAGGCCCTGG	891310	61.68	855
	IR-r start (pb)	1541953	2-r	GAAGCCCAGCAGATCCTG	1541187	62.47	GGCCTTGGCGAATTCTCA	1541919	62.71	732
	IR-l end (pb)	1543305	2-l	GACAAACGACCTCCCAGAC	1543360	62.54	CCACATAGTTGAGATCGAAGG	1544015	60.25	655
Multiplex 2	IR-r start (pb)	1987704	3-r	TTGAACGTTGACGTGATGTC	1987052	61.18	TCAAACAGTCAGCCGATCC	1987607	62.26	555
	IR-l end (pb)	1989056	3-l	CCGTGATGGGCTTTGTTAC	1989252	61.33	CACTCCAAGTCACAGCTCA	1989695	61.81	443
	IR-r start (pb)	1996102	4-r	CATGCAGCAATGCGAACTC	1995739	62.70	CCGACCTTTGTCTGTTGA	1996094	62.12	355
	IR-l end (pb)	1997454	4-l	GCAACCGACGATTGATCTGA	1997620	62.89	AATGTTGAGGCGTTTCTGC	1997875	61.56	255
Multiplex 3	IR-r start (pb)	2365415	5-r	TTGCTATGCGCGCTTACTG	2365127	62.70	AAAGTGCTCGACTTGGTGG	2365287	62.78	160
	IR-l end (pb)	2366767	5-l	CACGTTCCGTTATTCGGTGA	2366916	62.25	ACTCCCTCTTGAACGCTTAG	2366971	61.58	55
	IR-r start (pb)	2430118	6-r	GATGGTCGCGTCGAGAAG	2429131	62.83	ATATTGAGTCCGGTTTGAACC	2430086	62.54	955
	IR-l end (pb)	2431470	6-l	CTTCCCGAGACTCGSATTG	2431475	62.73	CGTCGACATACACCGTCAG	2432312	62.68	837
Multiplex 4	IR-r start (pb)	2550015	7-r	CGGTAGTCGTCGACAGTTT	2549196	61.91	AGACTGACGACCCGATGAT	2549949	62.67	753
	IR-l end (pb)	2551367	7-l	GTTGGAAAGTGAGGCCTCC	2551539	62.88	ATGCAGCTTGACGATCACC	2552194	62.99	655
	IR-r start (pb)	2635578	8-r	TTGATGTTGCCGAGTTCC	2634921	62.87	GCGACACGAACTTCGGTAA	2635476	62.85	555
	IR-l end (pb)	2636930	8-l	ACAACCTCGGTTTCTGCAGG	2637336	62.78	GGTATCTCCAACCTCGGGTG	2637791	61.73	455
Multiplex 5	IR-r start (pb)	2784616	9-r	CCTAAGCCGATATGGGCAG	2784121	62.26	TATCACCGTCAATTGCTGGG	2784476	62.59	355
	IR-l end (pb)	2785968	9-l	AATGAAGCGGTAACGTGGAT	2786132	62.14	CAGTAACTATCGGCGGAG	2786387	62.83	255
	IR-r start (pb)	2972110	10-r	TTAGGCTCCATGGAGCCTAA	2971853	62.60	TAAACCATCCGGTAGCCA	2972008	62.96	155
	IR-l end (pb)	2973462	10-l	GTAGTCGTTGACTGATCTCGC	2973609	62.85	AGAACCAAGAGGGTCAAGAG	2973664	61.16	55
Multiplex 6	IR-r start (pb)	3551231	12-r	TTACTGTTGGCGAAGCG	3550354	62.38	ATTGGGAACACGAGCTGTG	3551111	62.86	757
	IR-l end (pb)	3554066	13-l	CTTGTGGACCTACAGCAC	3554148	62.56	TTCGATGCGTACCAGACTC	3554804	61.71	656
	IR-r start (pb)	3710383	14-r	ATATCCGTGATCGTCATGCG	3709814	62.21	TGTCATAACAAGCGGTGCG	3710379	62.6	565
	IR-l end (pb)	3711735	14-l	CACAAGTACAGCCACCTT	3711772	62.12	GCTCGGATCGAGTTTCATG	3712225	62.81	453
Multiplex 7	IR-r start (pb)	3795059	15-r	TCATTGAGCACGACGATCAC	3794547	62.88	CATTGCATCGGGTGTTCAC	3794902	61.99	355
	IR-l end (pb)	3796411	15-l	CTATCGAGCGTTCACCAC	3796497	62.72	GTCGGTGAATGTGACGTTG	3796752	61.60	255
	IR-r start (pb)	3890780	16-r	GAAGCTGACCAACCGATAGTC	3890577	62.84	GTGGTGTCTTGCATGCC	3890732	61.96	155
	IR-l end (pb)	3892132	16-l	TTTCGCAGTGATTGCAAG	3892321	61.62	CGCATAGCTGTAGCGCTCT	3892376	61.87	55

Tableau 4. Description des amorces utilisées lors de l'amplification des séquences adjacentes à IS6110, IR-r, (pour Inverted Repeat Right), fait référence à la séquence inversée répétée qui se situe à l'extrémité 5' de l'IS6110, IR-l, (pour Inverted Repeat Left), fait référence à la séquence inversée répétée qui se situe à l'extrémité 3' de l'IS6110, Le nom de l'amplicon comprend l'identifiant (ID) de l'IS6110 suivi de la lettre "r" (right) pour les fragments situés en 5' ou "l" (left) pour ceux en 3' par rapport à l'IS6110,

Dans le génome d'une souche H37Rv, (ayant pour numéro d'accès NC_000962 sur le serveur NCBI), 16 copies d'IS6110 ont été répertoriées. L'IS6110 ayant pour

identifiant ID11 n'a pas été retenue, vu qu'elle se trouve dans le locus DR (tableau 4). Le locus DR est déjà connu pour être sujet à des pertes génomiques, il est donc inutile de cibler des fragments d'ADN adjacents à cette IS6110, d'autant plus qu'on ne pouvait pas s'attendre à avoir des amplicons de taille fixe. L'IS6110AD-typing consiste donc en l'amplification de 28 séquences génomiques qui encadrent les différentes IS6110 considérée. Deux copies d'IS6110 sont contiguës parmi les 15. Il n'y a pas d'amplification entre ces deux IS6110, ce qui ramène le nombre total de fragments amplifiés à 28 au lieu de 30,

La position des 15 copies d'IS6110 considérées chez H37Rv dans l'IS6110AD-typing, ainsi que les 28 amorces utilisées pour amplifier les séquences environnantes, sont décrites dans le tableau 4.

1.2.4.2. Description du protocole

Le protocole commence par l'exécution d'une PCR. Le Mix est préparé en suivant les indications du tableau 5. Le programme enregistré dans le thermocycleur doit correspondre à celui indiqué dans le tableau 6. Après la PCR, les produits PCR sont séparés par électrophorèse sur un gel d'agarose (2,5%). La taille des fragments est déduite par comparaison grâce à la migration d'un marqueur de poids moléculaires de 50-bp.

		Concentration (multiplex pour 4 amplifications)	Concentration (multiplex pour 5 amplifications)	Volume μ l (multiplex pour 4 amplifications)	Volume μ l (multiplex pour 5 amplifications)
Buffer (X)	10	10	10	2.1	2.1
MgCl ₂ (mM)	25	25	25	1.68	1.68
Paire d' amorces 1	Amorces F (μ M)	16	20	0.394	0.315
	Amorces R (μ M)	16	20	0.394	0.315
Paire d' amorces 2	Amorces F (μ M)	16	20	0.394	0.315
	Amorces R (μ M)	16	20	0.394	0.315
Paire d' amorces 3	Amorces F (μ M)	16	20	0.394	0.315
	Amorces R (μ M)	16	20	0.394	0.315
Paire d' amorces 4	Amorces F (μ M)	16	20	0.394	0.315
	Amorces R (μ M)	16	20	0.394	0.315
Paire d' amorces 5	Amorces F (μ M)		20		0.315
	Amorces R (μ M)		20		0.315
DMSO (%)	100	100	100	2.1	2.1
dNTPs (mM)	25	25	25	0.42	0.42
rTAQ U/ μ l	5	5	5	0.084	0.084
ADN (μ g/ μ l)	0.1	0.1	0.1	1.5	1.5

Tableau 5. Composition du Mix pour chaque multiplex de l'IS6110AD-typing.

Cycle	Température (°C)	Temps
1	94	10 min
2*	94	30 secs
3*	58	1 min
4*	72	2 min
5	72	10 min

Tableau 6. Description des cycles de la PCR de l'IS6110AD-typing. Le 2^{ième}, 3^{ième} et 4^{ième} cycle (2*, 3*, 4*) sont répétés 35 fois.

1.2.5. Le génotypage basé sur le polymorphisme de longues séquences (LSP)

Les LSP ont déjà été décrits dans le chapitre « Introduction ». La mise en évidence de la présence ou de l'absence de ces longues séquences, passe par une PCR avec les couples d'amorces spécifiques à la région ciblée. La révélation des fragments se fait après une électrophorèse.

1.3. PCR en temps réel et SNPs (Polymorphismes de nucléotide unique)

Le but de cette méthode est de quantifier la fluorescence émise par le milieu réactionnel durant l'amplification de l'ADN. Cette technique repose sur 2 bases importantes : soit on observe les 2 brins d'ADN en utilisant un intercalant fluorescent (ex. SYBR Green), soit on considère 1 brin en utilisant des sondes fluorescentes. Dans ce dernier principe, l'une des technologies les plus utilisées est l'hydrolyse de sondes *Taqman*.

Nous disposons d'une sonde fluorescente propre à sa séquence cible se situant à l'intérieur de l'amplicon et d'amorces spécifiques pour générer ces amplicons. Aux extrémités de ces sondes sont fixées, un fluorochrome émetteur (ex. FAM : 6-carboxyfluorocein) en 5' et un fluorochrome suppresseur à l'extrémité 3' (ex. TAMRA : 6-carboxy-tetramethyl-rhodamine). Dans cette configuration de fluorochromes proches, l'émission de fluorescence par FAM est inhibée (absorbée) par TAMRA. Lors de l'hybridation, la sonde et les amorces se fixent à leurs séquences cibles. A ce stade, il n'y pas de fluorescence détectable. Durant l'élongation, il est important que les températures soient de telle sorte que la sonde reste hybridée à sa cible. Pour se faire, le T_m (température de dissociation) de la sonde, proche de 70°C, doit être plus élevée de (5 à 10°C) que le T_m des amorces.

Ainsi la sonde reste stable durant l'élongation (à 60°C). La polymérase peut donc débiter l'élongation du nouveau brin d'ADN à partir de l'amorce jusqu'à la sonde (si hybridée), qu'elle hydrolysera (activité 5'-exonucléase). La sonde étant hydrolysée, le reporter (non lié à son suppresseur) émet de la fluorescence qui augmentera à chaque cycle proportionnellement au taux d'hydrolyse de la sonde (figure 5).

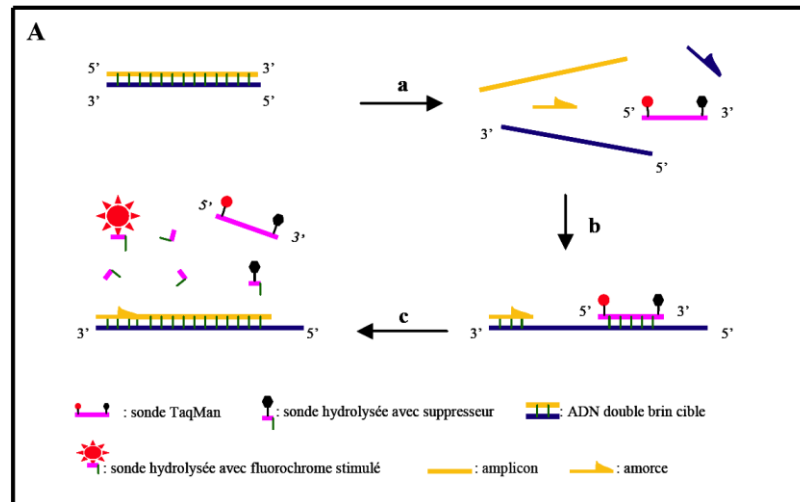


Figure 5. Hydrolyse de sondes (*Taqman* assay) (a) Durant l'étape de dénaturation, la sonde est libre en solution. (b) À la température d'appariement, la sonde et les amorces s'hybrident à leurs séquences cibles et la proximité des fluorochromes permet l'inhibition de la fluorescence. (c) La polymérase déplace et hydrolyse la sonde. Le fluorochrome émetteur est libéré de l'environnement du suppresseur permettant ainsi l'émission de la fluorescence. (Poitras et Houde, 2002).

Protocole PCR. Le typage des SNPs par PCR en temps réel suit plusieurs étapes :

- Le mélange des différents réactifs utiles à la PCR est résumé dans le tableau 7,
- La programmation du thermocycleur en temps réel suit les cycles du tableau 8.

Le principe du « cycle seuil » (ou Threshold cycle, Ct) est à la base des techniques fluorescentes en PCR. Le « cycle seuil » correspond au nombre de cycles à partir duquel le produit PCR est détectable et dépend fortement de la quantité d'ADN initiale. Au cours de chaque cycle la fluorescence augmente et correspond à la quantité d'amplicons produits à un moment donné (figure 6).

Réactifs	Concentration Solution-mère μM	Concentration Finale nM	Volume pour 1 Reaction (μl)
MasterMix 2 x	2 x	1 x	12,50
Primer forward	20	900	1,13
Primer reverse	20	900	1,13
Sonde TaqMan 1	20	200	0,25
Sonde TaqMan 2	20	200	0,25
H ₂ O			4,75
Volume total (μl)			20,00
Ajouter	5	μl ADN à	20

Tableau 7. Préparation des différents réactifs composant le Master Mix pour PCR en temps réel.

Température	Durée	Répétition
50 °C	2 min	1
95 °C	10 min	1
92 °C	15 secs	35
60 °C	1 min	

Tableau 8. Description des cycles de températures de la PCR en temps réel.

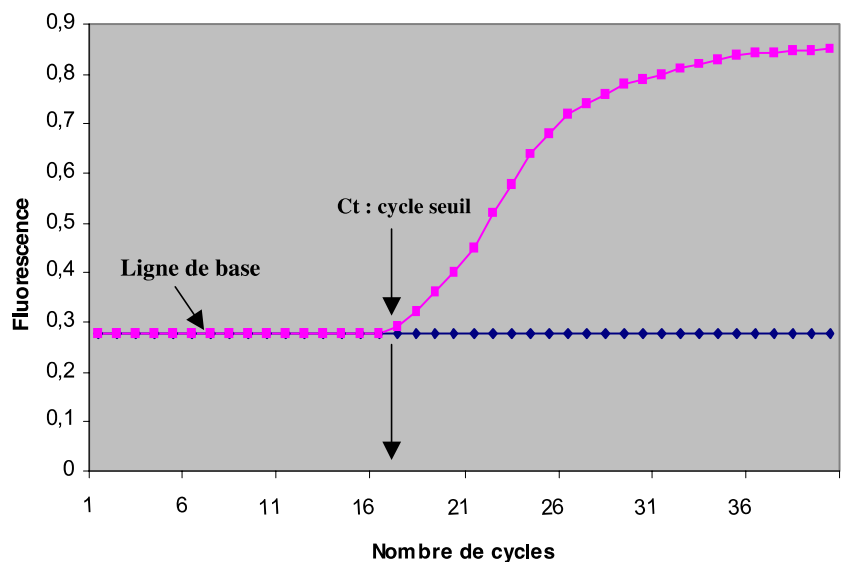


Figure 6. Exemple de courbe d'intensité de fluorescence exprimée en fonction du nombre de cycles. L'intensité de la fluorescence à chaque cycle est proportionnelle à la concentration d'amplicons, le cycle seuil (Ct) représente le nombre de cycles requis où le signal d'émission de fluorescence est statistiquement et significativement plus élevé que la ligne de base. (Poitras et Houde, 2002)

2. SITVITWEB : collection génotypique internationale de MTC

2.1. Généralités sur la collection de données de SITVITWEB

La tuberculose (TB) reste une cause majeure de morbidité et de mortalité dans le monde entier. En plus du diagnostic rapide et de la prise en charge des maladies et des patients, le suivi et l'établissement de rapports épidémiologiques constituent des points clés dans les stratégies lancées par l'OMS. Dans ce contexte (largement décrit dans l'introduction), le développement d'outils contribuant à l'identification et à l'analyse des différents clones appartenant au complexe *Mycobacterium tuberculosis* (MTC) qui circulent dans le monde est essentiel. Parmi les différentes méthodes de génotypage utilisées pour l'étude du polymorphisme génotypique de MTC, le spoligotypage et le typage basé sur les minisatellites MIRU-VNTR ont l'approbation internationale puisqu'ils se sont révélés être des marqueurs solides, rapides, reproductibles et surtout pratiques, vu qu'ils produisent des données dans un format portable. Le spoligotypage (Kamerbeek et al., 1997) constitue le pilier de la base de données en ligne SpolDB4 publié en 2006. Cependant, seule, cette méthode présente un faible pouvoir discriminatoire, de ce fait il est recommandé de l'utiliser de façon conjuguée avec les loci MIRU-VNTR (Supply et al., 2001, 2006) pour atteindre une bonne discrimination épidémiologique. Ce chapitre va donc décrire la base de données publique SITVITWEB, qui est une mise à jour de SpolDB4. Il intègre des données issues de marqueurs différents pour offrir une vision globale de la diversité génétiques de MTC dans le monde entier.

2.2. Description des diverses informations stockées dans SITVITWEB

SITVITWEB est la plus grande base de données de profils génotypiques de MTC. Elle compte 62 592 isolats cliniques décrits selon 3 marqueurs génotypiques différents (Spoligotype-43, MIRU12 et VNTR) et originaires de 153 pays différents. Cette nouvelle version publiée, donne accès en ligne à l'ensemble des souches collectées, avant septembre 2007, au sein du laboratoire des mycobactéries de l'Institut Pasteur de la Guadeloupe. Ces souches ont été isolées par divers investigateurs sur un total de 105 pays. Chaque entrée est étiquetée par un identifiant encore appelé « Isonumber » qui obéit à un formalisme précis (code pays + code laboratoire d'isolement + année d'isolement + code sensibilité aux antibiotiques + nom de la souche). Au-delà des informations de typage, chaque isolat clinique est décrit par son pays d'origine et d'isolement, sa ville et son année

d'isolement, son investigateur, ainsi qu'un code de laboratoire d'isolement et un code de résistance aux antibiotiques.

Dans la description de SITVITWEB, certains termes seront récurrents, donc il convient ici de les définir. Un « pattern » désigne un résultat de typage quelle que soit la méthode. Le pattern désigne donc un profil génotypique. Un « International Type » ou « IT » est un chiffre attribué à un pattern lorsqu'il est observé plus d'une fois dans la base de données. Ainsi SIT (Spoligotype International Type), MIT (MIRU International Type) ou VIT (VNTR International Type) font référence respectivement à un pattern de spoligotype43, de MIRU12 (pour les 12 loci MIRU-VNTR) et de VNTR (pour les 5 loci ETR) présent au moins deux fois dans la base de données, par opposition aux patterns orphelins qui sont uniques. Le terme « pattern endémique » désigne un pattern présentant une spécificité géographique, par opposition à « ubiquitaire », qui qualifie celui qui est retrouvé partout. Clade et lignée sont des synonymes qui font allusion à la classification des souches MTC selon le schéma de classification spoligotypique (Brudey et al., 2006) ou celui basé sur les LSP (Gagneux et al., 2007).

Pour les analyses géographiques, 6 régions et 23 sous-régions furent considérées. Cette subdivision géographique suit celle établie par l'organisation internationale des Nations Unies (UN). L'UN divise le monde en 5 régions et 22 sous-régions. Pour mieux satisfaire les besoins des études épidémiologiques, la région Antarctique fut ajoutée ainsi que la sous-région Asie du Nord (comprenant seulement la Russie). La région Amérique compte 4 sous-régions ; l'Amérique du Nord (AMER-N), centrale (AMER-C), du Sud (AMER-S) et les Caraïbes (CARI). L'Europe compte également 4 sous-régions ; l'Europe de Nord (EURO-N), du Sud (EURO-S), de l'Est (EURO-E) et de l'Ouest (EURO-W). L'Afrique regroupe 5 sous-régions ; l'Afrique du Nord (AFRI-N), centrale (AFRI-M), du Sud (AFRI-S), de l'Est (AFRI-E) et de l'Ouest (AFRI-W). L'Asie rassemble 6 sous-régions ; l'Asie du Nord (ASIA-N), du Sud (ASIA-S), centrale (ASIA-C), du Sud-Est (ASIA-SE), de l'Est (ASIA-E), et de l'Ouest (ASIA-W). L'Océanie compte 4 sous-régions géographiques. L'Australie et la Nouvelle-Zélande constituent une sous-région. La Mélanésie, la Micronésie et la Polynésie sont les 3 dernières.

2.3. Comparaison avec les précédentes versions de la base ; SpoIDB1, SpoIDB2, SpoIDB3 et SpoIDB4

Les précédentes versions de la base de données SpoIDB1, SpoIDB2, SpoIDB3 et SpoIDB4 ne présentaient que des données de spoligotypage. La première version de la base de données remonte à 1999 avec SpoIDB1 (Sola et al., 1999). Celle-ci contenait 610 isolats cliniques, 69 SITs et aucun orphelin. La seconde version de la base de données mondiale de spoligotypes, SpoIDB2 (Sola et al., 2001), comptait 3 319 isolats avec 259 SITs et 540 patterns orphelins. La version qui a suivie, SpoIDB3 (Filliol et al., 2002) comptait 11708 isolats cliniques qui provenaient de 90 pays différents. Elle comptait 813 SITs et 1 300 patterns orphelins. SpoIDB4 (Brudey et al., 2006), la dernière version à être publiée, donnait des informations de spoligotypage sur 39 295 souches provenant de 122 pays. Elle regroupait 1 939 SITs et 3370 patterns orphelins. SITVITWEB, à son tour, compte 53 816 isolats décrits avec le spoligotypage. La base atteint 2 740 SITs et 4 364 patterns orphelins (tableau 9).

Base de données	Marqueur moléculaire	Nombre de patterns	Nombre de souches	Nombre de IT	Nombre de souches avec un IT	Nombre de souches avec un pattern orphelin	Pourcentage de souches avec un pattern	Nombre de pays	HGDI
SpoIDB1	Spoligotype		610	69	610	0	0.0	> 5	
SpoIDB2	Spoligotype	799	3 319	259	2 779	540	16.3	47	0.930
SpoIDB3	Spoligotype		13 008	813	11 708	1 300	10.0	> 90	0.974
SpoIDB4	Spoligotype	5 309	39 295	1 939	35 925	3 370	8.6	122	0.980
SITVITWEB	Spoligotype	7 104	58 187	2 747	53 830	4 357	7.5	102	0.979
	MIRU12	2 380	8 161	847	6 628	1 533	18.8	35	0.995
	VNTR (ETR)	458	4 626	245	4 413	213	4.6	22	0.945

Tableau 9. Composition des différentes versions de la base de données génotypique de MTC

Un réel effort a été porté au nettoyage des données, sur SITVITWEB. Avant leur intégration, les données sont comparées à la base pour éviter d'éventuelles duplications. Lorsqu'un orphelin présente un pattern jugé suspect, l'investigateur est contacté pour une vérification. Ainsi, près de 10 000 souches ont subi des corrections relatives à leurs informations de typage. Comme le montre la Figure 7, le nombre de souches SITées n'a pas cessé de croître. Cependant, il a y une décroissance du pourcentage de souches présentant un pattern orphelin. On arrive donc à une conclusion logique : plus on accumule des données, plus il est probable que les nouvelles données soient des patterns déjà observés. Cela est la preuve que la diversité spoligotypique de SITVITWEB tend à être exhaustive.

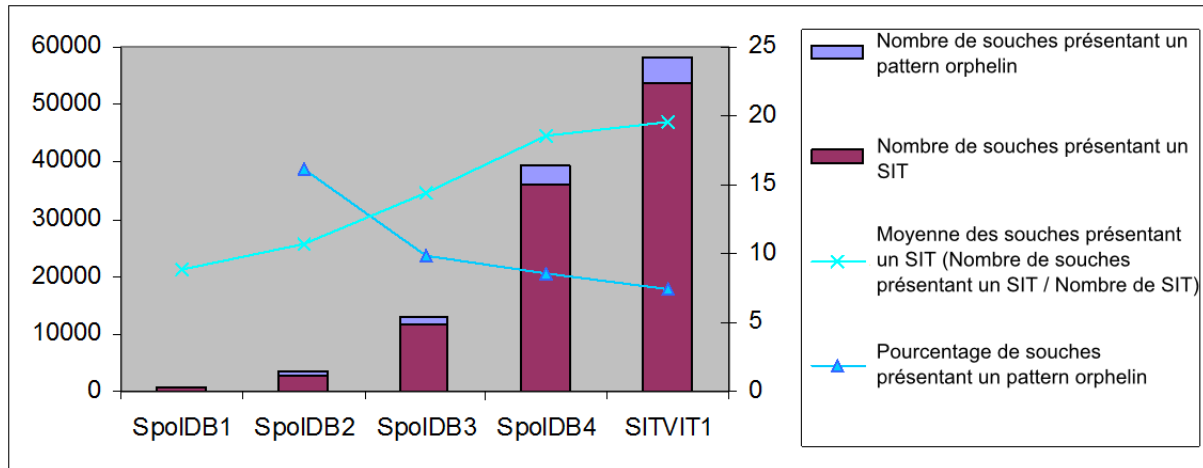


Figure 7. Evolution des différentes versions de la base de données génotypique de MTC

3. Les logiciels de phylogénie utilisés

3.1. MrBayes pour une phylogénétique selon la méthode bayésienne

MrBayes3 (Ronquist et al. 2003) (<http://mrbayes.csit.fsu.edu/>) est utilisé pour faire les relations phylogénétiques entre des isolats de MTC. Pour commencer, il faut spécifier le taux de mutations. Pour les loci MIRU-VNTR, vu que ce taux varie d'un locus à un autre, la distribution Gamma semble appropriée. Ensuite, il faut préciser les paramètres du processus de Monte-Carlo par chaîne de Markov, c'est-à-dire le nombre de générations et la fréquence d'échantillonnage. Par exemple, pour l'arbre bayésien de la figure 22 du chapitre « Résultats et Discussion », il a fallu 13 millions de générations, et sauvegarder un arbre toutes les 10 mille générations. Cela a donc aboutit à 1 300 échantillons provenant de la distribution de probabilité a posteriori. Ensuite, 25% des premiers arbres c'est-à-dire les 325 premiers arbres sont supprimés car ils font partie de l'étape d'approche.

Enfin, l'arbre de la figure 22 est un consensus majoritaire des 975 arbres restants. Cet arbre consensus a été généré avec le logiciel Dendroscope (<http://ab.inf.uni-tuebingen.de/software/dendroscope/>).

3.2. BioNumerics pour la construction phylogénétique selon la méthode du Minimum Spanning Tree (MST)

BioNumerics (version3.5, Applied Maths, Sint-Marteen-Latem, Belgium) a été utilisé pour son algorithme basé sur le "Minimum Spanning Tree" (MST). Cette méthodologie est idéale pour montrer comment les patterns VNTR d'une même

lignée s'agglomèrent en étoile autour d'un pattern central. Le MST a été utilisé plusieurs fois dans le cadre de cette thèse, pour compléter, dans certains cas, un échantillon qui montrerait un défaut d'exhaustivité.

4. Outils mathématiques utilisés

4.1. Les réseaux bayésiens

4.1.1. Définition

Un réseau bayésien est composé :

- d'un graphe causal représentant les connaissances que l'on a d'un système informatique, biologique, sociologique etc. Ce graphe est constitué d'arcs illustrant les relations de causalité existant entre les variables (ou nœuds) qui décrivent le système concerné.
- de l'ensemble des distributions locales de probabilités utile au calcul, au sein du modèle, des probabilités conditionnelles d'évènements reliés les uns aux autres par les relations de cause à effet.

Ainsi avec ces deux composants, tous les problèmes d'inférence sont ramenés à des problèmes de théorie des graphes.

4.1.2. Fonctionnement

Un réseau bayésien est construit en réponse à un problème donné. Il combine la connaissance d'experts avec celle extraite des données. Il s'agit ici que l'expert représente graphiquement de la façon la plus intuitive possible ses connaissances, détermine les relations de dépendance entre variables.

L'utilisation des réseaux bayésiens repose sur la propagation de l'information dans les réseaux, autrement dit sur le calcul de la probabilité d'un évènement connaissant certaines observations. C'est ce que l'on appelle l'inférence. L'inférence permet le traitement de requêtes complexes. On peut vérifier la valeur la plus probable d'un nœud, la cause la plus probable d'une information, on peut évaluer la sensibilité à une information en mesurant les variations de la probabilité de cette hypothèse quand on fait une observation.

4.1.3. Choix de GeNIe pour l'apprentissage de structure et l'apprentissage des tables de probabilités

Le modèle bayésien pour *M.tuberculosis*, VNTR-TBbayes, a été construit à partir d'un échantillon de 3818 souches (de SITVIT2) décrit avec 57 variables (décrites dans la partie « Discussion et Résultats »). Le logiciel GeNIe (Graphical Network Interface, disponible en ligne <http://genie.sis.pitt.edu/about.html>), a été utilisé pour l'apprentissage de structure afin d'établir les relations de dépendance entre les variables. Le modèle fut construit avec l'algorithme "Greedy Thick Thinning", qui est basé sur une distribution du khi2 et implémenté selon la stratégie Greedy.

Un algorithme glouton (ou stratégie Greedy), est un algorithme, qui, à chaque étape fait un choix optimal localement dans l'espoir d'obtenir un résultat optimum global. Une structure Khi2 sélectionne le réseau le plus probable, "a posteriori", à partir d'un ensemble de réseaux possibles $P(B_i | D) / (B_j | D)$ (où B_i et B_j sont deux structures, alors que D est la base de données) dans le but d'ordonner un ensemble de modèles.

Le logiciel GeNIe a aussi fait l'apprentissage des tables de probabilités, ce qui a permis d'avoir un modèle fonctionnel.

Une fois le modèle construit, il a été testé sur un échantillon de 880 isolats provenant de SITVIT2.

4.2. La loi normale

4.2.1. Généralités

La distribution normale a tout d'abord été décrite par Abraham de Moivre (1667-1754) comme étant la limite de la loi binomiale pour n infiniment grand et $p = q = \frac{1}{2}$. Francis Galton (1822-1911) pense que souvent les processus naturels, (particulièrement en physique ou en biologie), obéissent à une distribution normale. Le mathématicien français Emile Borel (1877-1956) apporta les explications nécessaires à la réalisation d'une distribution normale. Selon lui, pour qu'une variable aléatoire réelle X suive une loi normale, il lui faut être sous la dépendance d'une cause majeure et d'un ensemble d'autres causes moins importantes ou secondaires et indépendantes. Les effets de ces paramètres sont additifs et permettent d'approcher la loi normale. Plus il y a des causes secondaires, plus l'approximation

est bonne. La distribution normale est un idéal qui ne se rencontre jamais dans la nature. Au mieux on observe une bonne approximation qui se traduit par une courbe en forme de cloche. Cette forme s'explique par le fait que la plupart des individus sont autour de la moyenne, et plus on s'écarte de cette moyenne moins on compte d'individus (symétriquement). Plus l'échantillon est grand, plus l'approximation à la distribution normale est bonne.

Une variable aléatoire χ suit une loi normale de paramètres m (pour la moyenne) et σ (pour l'écart type), notée $N(m,\sigma)$, lorsqu'elle a pour densité de probabilité sur \mathbb{R} la fonction :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

Cette fonction de densité mène à la courbe de Gauss de la figure 8.

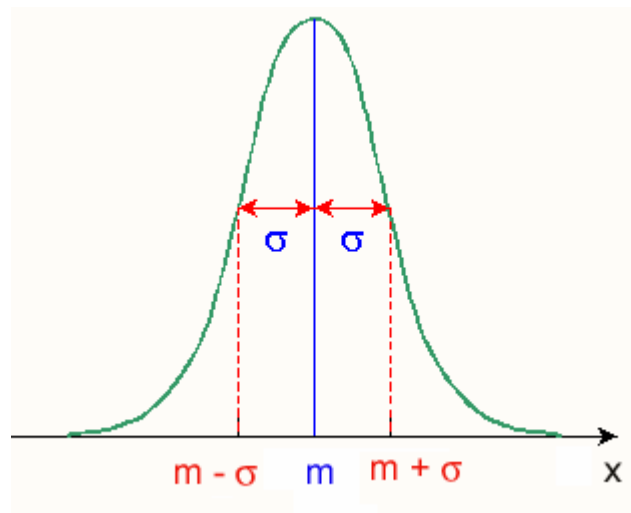


Figure 8. Illustration de la fonction de Gauss. Image tirée du document : http://ufr-pluribab.univ-pau.fr/live/digitalAssets/11/11523_L2_loi_normale.pdf

4.2.2. Renforcement de la prédiction du nombre de répétitions portées par chaque locus MIRU-VNTR grâce au Z-score

La loi normale est introduite dans ce chapitre car pour chaque locus MIRU-VNTR, dans chaque lignée, la distribution du nombre de répétitions se superpose à une distribution gaussienne.

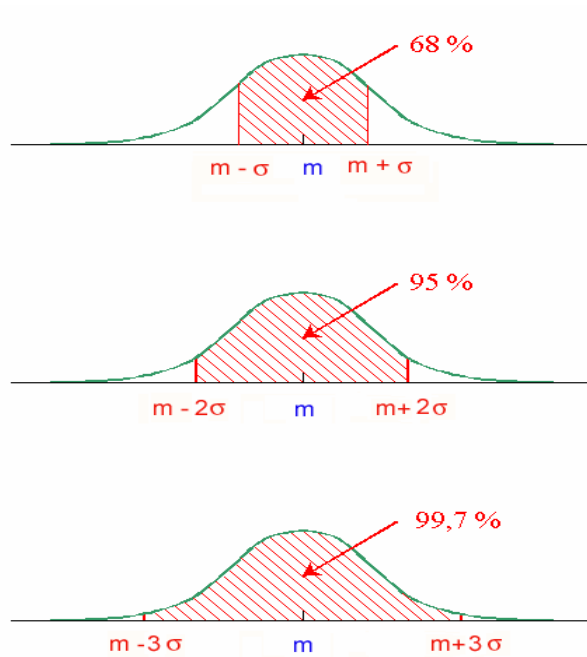


Figure 9. Illustration de quelques propriétés de la distribution gaussienne. Image tirée du document : http://ufr-pluribab.univ-pau.fr/live/digitalAssets/11/11523_L2_loi_normale.pdf

Il est admis que pour une variable qui suit une loi normale :

- 68.3% des valeurs qu'elle prend se situent dans l'intervalle $[m-\sigma, m+\sigma]$ (figure 9),
- 95.5% des valeurs qu'elle prend se situent dans l'intervalle $[m-2\sigma, m+2\sigma]$ (figure 9),
- 99.7% des valeurs qu'elle prend se situent dans l'intervalle $[m-3\sigma, m+3\sigma]$ (figure 9).

Pour fixer un intervalle de confiance à la plage de valeurs possibles proposées par VNTR-TBbayes pour un locus sachant d'autres variables du système, le Z-score a été utilisé. Le Z-score se définit comme suit :

$$Z = \frac{x - m}{\sigma}$$

(où x = nombre de répétitions observé pour un locus, m = valeur moyenne de répétitions pour un locus, et σ = écart type). Si le score est compris entre -2 et 2, le nombre de répétitions x a 95.5% de chance de retrouver dans l'intervalle $[m-2\sigma, m+2\sigma]$.

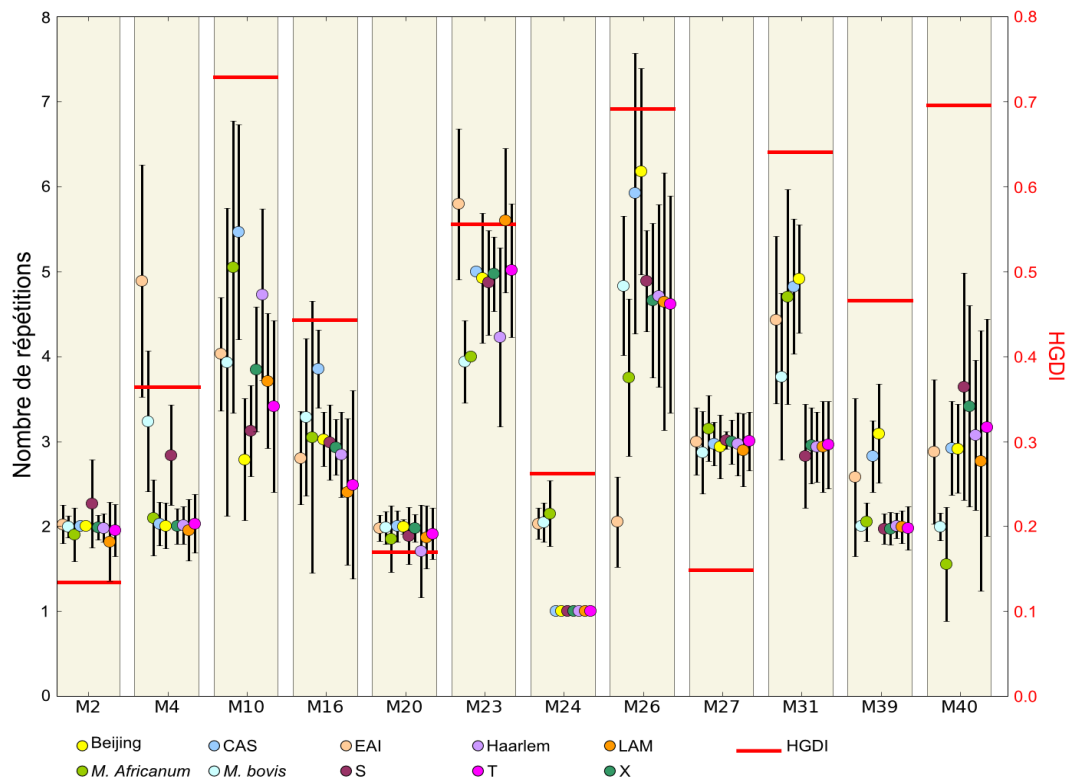


Figure 10. Les variables considérées pour le calcul du Z-score. La moyenne et l'écart type ont été calculés pour chaque locus au sein de chaque lignée (Beijing, CAS, EAI, Haarlem, LAM, *M. africanum*, *M. bovis*, S, T, X). L'échantillon utilisé pour effectuer les calculs compte 3818 isolats du complexe *M. tuberculosis* provenant de SITVITWEB (Demay et al., 2012). Les moyennes sont symbolisées par des cercles et l'écart type par les barres d'erreur. L'HGDI global de chaque locus est représenté par une ligne horizontale rouge.

Afin de permettre le calcul du Z-score au sein de l'application bayésienne développée (Article 6), la moyenne et l'écart type de chaque locus à l'intérieur de chaque famille ont été calculés (figure 10). Ce Z-score renforce les prédictions faites par VNTR-TBbayes sur les loci MIRU-VNTR.

4.2.3. Utilisation du Z-score cumulé pour la classification des souches MTC selon leur valeur de MIRU12

Une classification basée sur MIRU-VNTR a été développée dans le cadre de cette thèse, classification totalement indépendante du schéma de classification spoligotypique. Dans la partie « Résultats et Discussion », il est démontré qu'un prototype MIRU-VNTR, est un pattern MIRU12 qui définit une lignée de MTC. Tous les variants d'une lignée se dispersent autour de ce pattern encore appelé nœud central. Ceci s'explique par le fait que la distribution des valeurs prises par chaque locus suit une loi normale. Après avoir identifié tous les nœuds centraux MIRU12 chez MTC, il a fallu développer un algorithme pour classifier un pattern quelconque

dans l'une des lignées nouvellement définies. L'algorithme de classification passe par deux étapes. Le pattern est comparé à l'ensemble des prototypes et retient les prototypes avec lequel il partage le plus grand nombre de valeurs de loci. Si un seul prototype est retenu, il dévoile la lignée d'appartenance du nouveau pattern. Si plus d'un prototype est retenu, l'algorithme passe à la seconde étape. Pour arriver à ses fins, l'algorithme de classification confronte le pattern à chaque prototype retenu. À chaque confrontation, il calcule un Z-score cumulé (ou cumulative Z-score, CZS), qui se définit comme suit :

$$CZS = \sum_{j=1}^n \left| \frac{x_j - X_j}{S_j} \right|$$

Le CZS se définit comme la somme de la valeur absolue de $\frac{x_j - X_j}{S_j}$ (qui correspond au Z-score) de chaque locus présentant des valeurs différentes. n est le nombre total de loci dissemblables, x_j est le nombre de répétition pour le $j^{\text{ième}}$ locus, X_j est la valeur moyenne de répétition du $j^{\text{ième}}$ locus et S_j est l'écart type du $j^{\text{ième}}$ locus. Le CZS entre un pattern et un prototype peut être traduit comme étant la distance qui les sépare. Ainsi, le prototype contre lequel le pattern présente le CZS la plus faible révèle la lignée d'appartenance.

4.3. L'index de discrimination de Hunter-Gaston (HGDI) pour évaluer la variabilité biologique

L'index de diversité Hunter-Gaston Diversity (HGDI), est couramment utilisé pour comparer les méthodes de typage en évaluant leur pouvoir discriminant propre, ou encore mesurer la variabilité d'un locus. La définition de l'HGDI dérive de celle de l'index de diversité de Simpson qui quantifie la biodiversité d'un habitat (Simpson, 1949). L'HGDI se définit comme suit :

$$HGDI = 1 - \frac{1}{N(N-1)} \sum_{j=1}^S n_j (n_j - 1)$$

Où N est l'effectif total de l'échantillon étudié, S est le nombre total de patterns différents dans l'échantillon ou de valeurs différentes pour un locus, n_j est le nombre de souches qui portent le $j^{\text{ième}}$ pattern ou valeur. La figure 9 donne l'HGDI de chaque locus MIRU-VNTR, et ce qu'importe la lignée spoligotypique.

RÉSULTATS ET DISCUSSION

1. Etat des lieux de la pertinence et des potentialités des données stockées dans la plus grande base de données génotypiques SITVITWEB (Article 1)

1.1. SITVITWEB : une mise à jour de la diversité génotypique des souches MTC circulant dans le monde

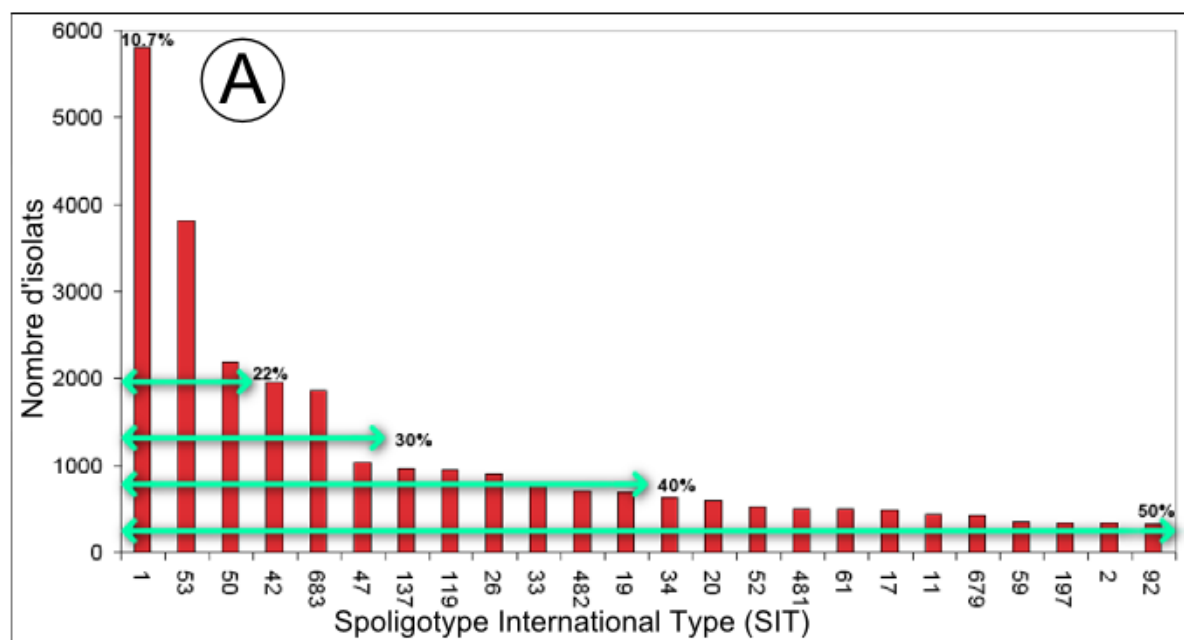
1.1.1. Apport des différents marqueurs moléculaires dans la description des profils génotypiques

Si SITVITWEB délivre des informations de génotypage de trois marqueurs moléculaires différents (spoligotype43, les 12 loci MIRU-VNTR ou MIRU12, les 5 loci ETR ou VNTR), ces trois types d'information ne sont pas disponibles pour tous les isolats répertoriés. Il est donc intéressant d'étudier la représentativité de chaque marqueur dans la base séparément. Dans ce but, les figures 1, 2 et 3 illustrent respectivement la distribution des SITs, MITs et VITs prédominants qui regroupent 50% de la totalité des isolats cliniques.

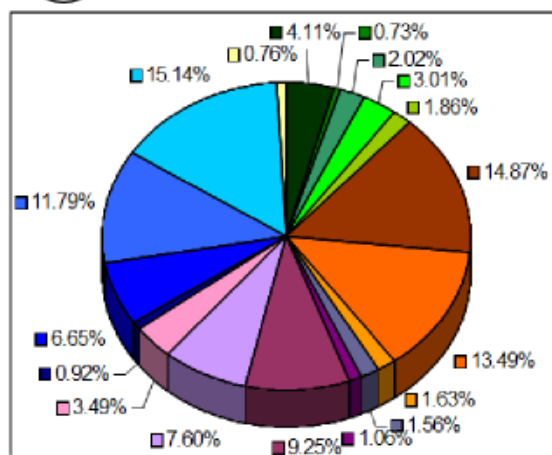
1.1.1.1. Apport du spoligotypage

L'information de spoligotypage est disponible pour 58 180 souches dans SITVITWEB. Sur un total de 7 104 patterns de spoligotype43, 2 740 sont des SITs. De ce fait, 4 364 patterns sont orphelins, ce qui représente 7% des souches de la base. Par rapport à SpoIDB4, SITVITWEB compte 801 nouveaux SITs, ainsi le pourcentage de souches SITées atteint 93% (n=53 816) des isolats cliniques de la base. Les 24 SITs prédominants englobent plus de la moitié de l'ensemble des isolats stockés (n=27 059) (figure 1A). Parmi les 24 SITs prédominants, 6 ne faisaient pas partie du top 24 de SpoIDB4. Seuls les trois premiers SITs (SIT1, SIT53, SIT50) ont maintenu leur position. On ne peut tirer des conclusions sur la surreprésentation de certains patterns du fait que toutes les régions n'ont pas la même représentativité dans la base. Par exemple 37,4% et 30,30% des isolats ayant un pattern SITé proviennent respectivement des régions Américaines et Européennes. L'Asie et

l'Afrique représentent tout de même respectivement 17,5% et 14,1% des souches SITées.



(B)



(C)

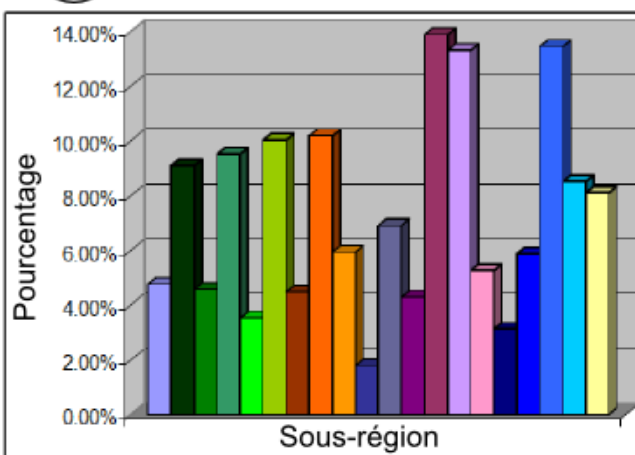


Figure 1. Description des patterns spoligotypiques. (A) Les 24 SITs prédominants classés par ordre décroissant du nombre d'isolats qu'ils regroupent dans SITVITWEB. Ces SITs englobent 50% des isolats cliniques de la base de données. Les flèches indiquent la proportion de souches que regroupent les SITs. (B) Le répartition des souches ayant un pattern orphelin dans les différentes sous-régions du monde. (C) Proportion de souches dans une sous-région donnée qui présentent un pattern orphelin par rapport au nombre total d'isolats cliniques répertoriés dans la sous-région.

Globalement, celles-ci couvrent 18 sous-régions géographiques dans le monde sur un total de 24. Seules 7,5% des souches sont orphelines. La plupart des souches présentant un pattern orphelin, se retrouvent en Europe et en Amérique avec respectivement 34,53% et 29,99% des orphelins de SITVITWEB (figure 1B). Le

pourcentage d'isolats ayant un pattern orphelin dans une région ne dépasse généralement pas 10% de l'ensemble des isolats. Cependant un peu plus de 12% des souches d'Asie du Sud, d'Asie du Sud-Est et d'Europe du Sud ont un pattern orphelin.

1.1.1.2. Apport du typage basé sur les loci MIRU-VNTR (formats 12 loci et 5 loci)

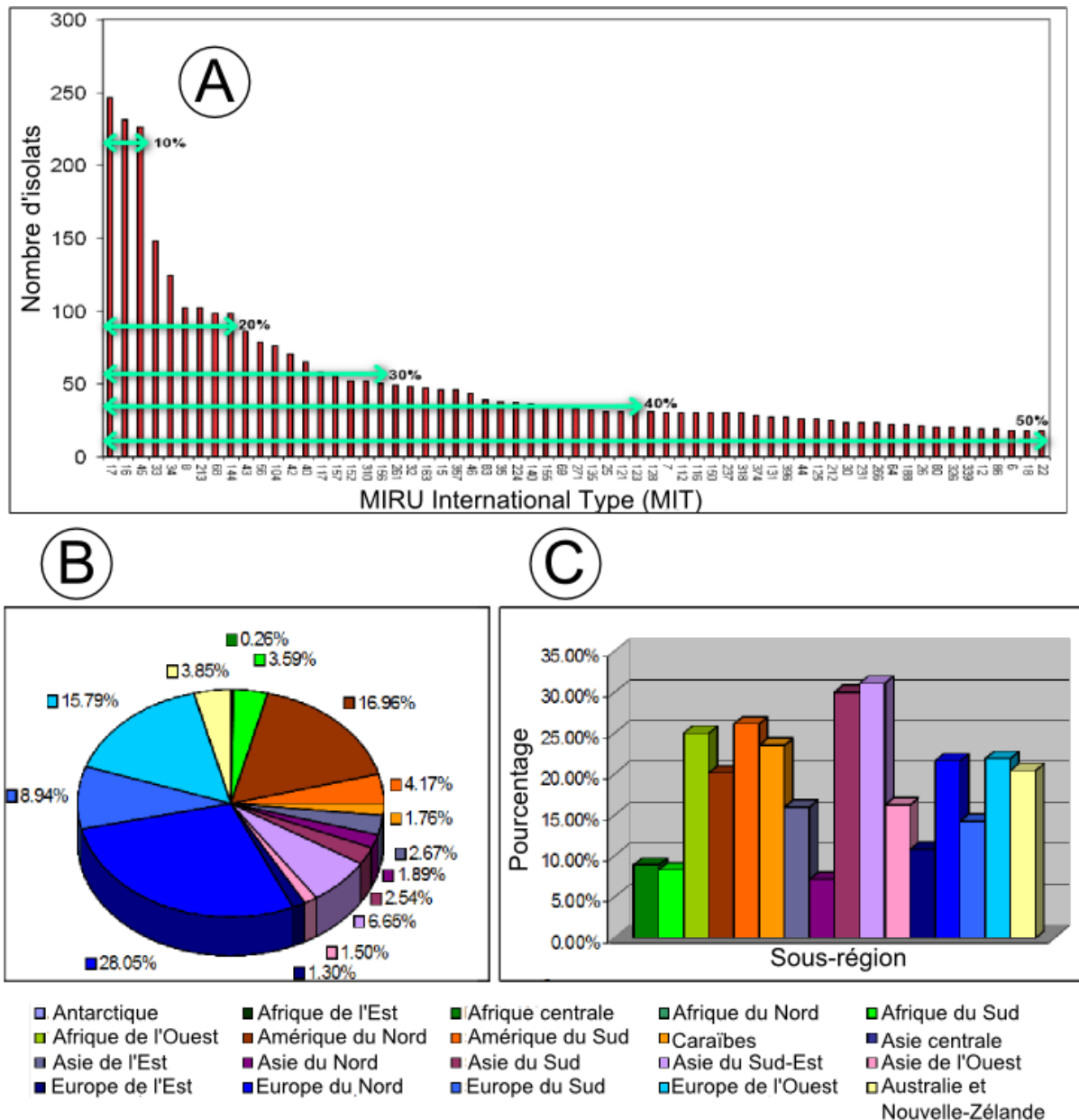


Figure 2. Description des patterns MIRU12. (A) Les 63 MITs prédominants classés par ordre décroissant du nombre d'isolats qu'ils regroupent dans SITVITWEB. Ces MITs englobent 50% des isolats cliniques de la base de données. Les flèches indiquent la proportion de souches que regroupent les MITs. (B) Le répartition des souches ayant un pattern orphelin dans les différentes sous-régions du monde. (C) Proportion de souches dans une sous-région donnée qui présentent un pattern orphelin par rapport au nombre total d'isolats cliniques répertoriés dans la sous-région.

La version de la base SITVITWEB, constitue la première intégration des loci MIRU-VNTR. 8 161 souches sont décrites avec les 12 MIRU-VNTR (MIRU12), ce qui représente 13% des souches de la base. Sur un total de 2 380 patterns MIRU12, 847 sont des MITs.

De ce fait, 1 533 patterns sont orphelins, ce qui représente 18,8% des souches décrites par les 12 loci MIRU-VNTR. Le pourcentage de souches MITées atteint 81,2% (n=6 628) des isolats cliniques de la base. 63 MITs englobent plus de la moitié de l'ensemble des isolats ayant un pattern MIRU12 (n=3329) (figure 2A). Les trois premiers MITs (MIT17, MIT16, MIT45) concentrent plus de 10% des isolats cliniques MITés. Comme pour les SITs, la proportion de souches MITées change d'une région à l'autre. L'Europe, à elle seule, regroupe 51,5% des souches MITées et 54% des souches avec un pattern orphelin (figure 2B). Globalement, la distribution des souches MITées couvrent 15 sous-régions géographiques dans le monde sur un total de 23 sous-régions. Le pourcentage d'isolats ayant un pattern orphelin dans une région ne dépasse généralement pas 25% de l'ensemble des isolats la région (figure 2C). Cependant, environ 30% et 12% des souches d'Asie du Sud et d'Asie du Sud-Est ont un pattern orphelin.

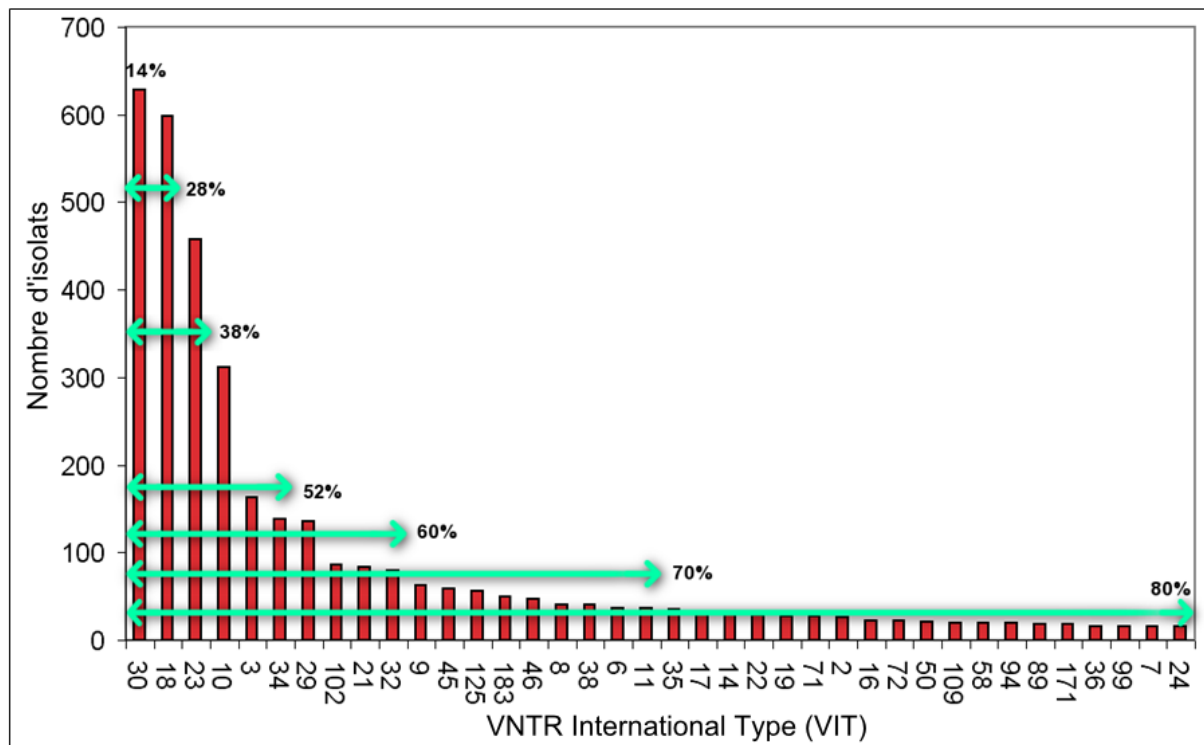


Figure 3. Illustration des 38 VITs prédominants classés par ordre décroissant du nombre d'isolats qu'ils regroupent dans SITVITWEB. Ces VITs englobent 80% des isolats cliniques de la base de données. Les flèches indiquent la proportion de souches que regroupent les VITs.

Dans SITVITWEB, le marqueur VNTR est également présent. Parmi les 62 592 isolats cliniques de SITVITWEB, 4 626 souches sont décrites avec les 5 loci ETRs (VNTR), ce qui représente 7,4% des souches de la base. Sur un total de 458 patterns VNTR, 245 sont des MITs et 213 sont orphelins. Les VITs rassemblent 4 413 isolats cliniques. La figure 3 illustre les 38 VITs prédominants qui regroupent 80% des souches VITées. Il est intéressant de remarquer que seuls 6 VITs (VIT30, VIT18, VIT23, VIT10, VIT3 et VIT34) rassemblent 50% des souches ayant un pattern VITé.

1.2. Intérêt du spoligotypage dans la description des profils génotypiques de MTC

1.2.1. Mise à jour des lignées spoligotypiques de SpolDB4

1.2.1.1. Méthodologie appliquée pour la mise à jour de la classification spoligotypique

Dans SpolDB4, 62 prototypes furent définis. La redéfinition de ces règles a requis plusieurs étapes. Tout d'abord l'ensemble des règles ont été appliquées à SITVITWEB. La plupart des souches furent bien classifiées. D'autres ont été mal classifiées ou non classifiées. Afin d'augmenter le nombre de souches classifiées, les règles ont été revues par un expert. Seules les souches classifiées furent considérées pour avoir une plus haute précision et fiabilité dans la définition des règles. De plus, les prototypes furent réévalués par un expert, dans un format binaire, alors que dans SpolDB4, leur définition était basée sur le format octal. Soixante sept prototypes ont pu être ainsi définis. Chacun est caractérisé par un ensemble d'espaceurs qui sont obligatoirement absents (représenté par des carrés rouges vides), des espaceurs qui sont obligatoirement présents (carrés rouges pleins), et d'autres espaceurs qui peuvent être absents ou présents (carrés noirs pleins) (figure 4).

1.2.1.2. Présentation des prototypes

Entre SpolDB4 et SITVITWEB, le nombre de prototypes est passé de 62 à 67 (figure 4). Afin de réduire le nombre d'isolats cliniques non classifiés, 5 règles ont été ajoutées. Il s'agit d'AFRI, de BOV, de CAS, de PINI et d'EAI2. Il est intéressant de remarquer que parmi les 24 SITs faisant partie du top 24 de la figure 1A, seuls 3

SITs (SIT481, SIT679, SIT197) ne font pas partie des prototypes. Ainsi, souvent, les prototypes, sont des patterns qui ont une forte occurrence dans SITVITWEB.

SIT Prototype	Clade Name	Binary Spoligotype Rules															Octal Code	Total	Frequency	New strains	Reclassified strains
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15					
181	AFRI1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77077777777671	379	0.74	109	0
331	AFRI2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77407760777071	81	0.16	7	0
438	AFRI3	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77407777777071	32	0.06	2	0
ND	AFRI4	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77477777777671	64	0.13	56	8
1	Beijing	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	000000000003771	6147	12.04	2096	6
482	BOV1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	676773777777600	2380	4.66	-21	2
683	BOV2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	664073777777600	2689	5.27	-81	0
479	BOV3	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	640013777777600	123	0.24	-81	0
647	BOV4-Caprae	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	277773777377600	279	0.55	113	2
ND	BOV5	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	676777777777600	690	1.35	604	86
592	Canettii	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	00000000101000	12	0.02	3	0
26	CAS1-Delhi	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	70377740003771	1490	2.92	384	0
21	CAS1-Kii	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	703377400001771	253	0.50	95	0
288	CAS2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	70037740003771	92	0.18	2	1
ND	CAS	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	70377740003771	473	0.93	468	5
48	EAI1-SOM	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	7777777413731	575	1.13	120	5
19	EAI2-Manila	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	67777747413771	842	1.65	305	2
89	EAI2-Nonthaburi	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	67400003413771	83	0.16	10	1
ND	EAI2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	67777747413771	4	0.01	4	0
11	EAI3-IND	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	4777777413071	566	1.11	133	1
139	EAI4-VNM	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777774413771	384	0.75	37	0
236	EAI5	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	7777777413771	1079	2.11	51	14
591	EAI6-BGD1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777757413771	331	0.65	154	8
1898	EAI7-BGD2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	7777770003371	70	0.14	2	0
109	EAI8-MDG	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	47777737413771	197	0.39	128	6
47	H1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	7777774020771	1827	3.58	414	6
2	H2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	00000004020771	349	0.68	53	0
50	H3	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	7777777720771	4287	8.40	2183	39
127	H4	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	7777747760771	166	0.33	-208	0
451	T (H37Rv)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	5777777420771	138	0.27	-11	2
20	LAM1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	677777607760771	813	1.59	310	19
17	LAM2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	677737607760771	677	1.33	265	11
33	LAM3	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	776177607760771	1539	3.02	1007	23
60	LAM4	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777607760731	372	0.73	213	11
93	LAM5	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777737607760771	437	0.86	166	8
64	LAM6	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777607560771	414	0.81	235	7
41	LAM7-TUR	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777404760771	353	0.69	191	2
290	LAM8	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777606760771	37	0.07	13	2
42	LAM9	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777607760771	2707	5.30	764	13
61	LAM10-CAM	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777743760771	620	1.21	187	1
59	LAM11-ZVE	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777606060771	697	1.37	457	0
209	LAM12-Madrid1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	776017607760771	31	0.06	5	2
100	MANU1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777777773771	88	0.17	13	0
54	MANU2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777777763771	243	0.48	83	0
1378	MANU3	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777777770771	104	0.20	73	3
539	Microti	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	00000000000600	29	0.06	-9	0
593	PINI1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	07400003777600	29	0.06	15	3
637	PINI2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	00000000777600	104	0.20	6	5
ND	PINI	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	07400003777600	5	0.01	3	2
34	S	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	776377777760771	1081	2.12	428	5
1737	T-Tuscany	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777760017760771	53	0.10	40	2
53	T1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777777760771	7779	15.24	1339	28
280	T1-RUS2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	770000777760771	81	0.16	16	2
52	T2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777777760731	947	1.86	113	2
135	T2-Uganda	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777777777760730	97	0.19	44	5
37	T3	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777737777760771	434	0.85	43	0
149	T3-ETH	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777000377760771	251	0.49	30	0
627	T3-OSA	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	741737777760771	29	0.06	9	1
40	T4	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777773777760771	149	0.29	-25	6
39	T4-CEU1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777347760471	189	0.37	55	1
44	T5	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777757760771	364	0.71	66	0
58	T5-Madrid2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777557760771	136	0.27	33	0
254	T5-RUS1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777760007760771	151	0.30	48	3
119	X1	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777776777760771	1256	2.46	299	1
137	X2	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	777776777760601	1431	2.80	236	1
92	X3	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	700076777760771	1137	2.23	49	4
405	ZERO	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	77777700000011	20	0.04	20	0

Figure 4. Description des 67 prototypes spoligotypiques. La 1^{ère} colonne indique le numéro de SIT du correspondant au prototype décrivant une ligne/sous-lignées, (ND pour Non Défini, terme choisi pour désigner le caractère atypique d'un pattern). La 2nd colonne indique le nom de la ligne/ sous-lignée. La 3^{ème} colonne donne la règle spoligotypique au format binaire (carrés pleins pour les espaceurs présents, carrés vides pour ceux qui sont absents, les carrés rouges désignent un état obligatoire, les carrés noirs désignent un état variable). La 4^{ème} colonne donne le code octal du pattern prototype. La 5^{ème} colonne indique le nombre d'isolats qui présentent le pattern. La 6^{ème} colonne donne la proportion de d'isolats qui présentent ce pattern par rapport au nombre total d'isolats SITés. La 7^{ème} colonne indique le nombre de nouvelles souches (par rapport à SpoDB4) qui présentent ce pattern prototype, (les valeurs négatives désignent le nombre d'isolats enlevés). La 8^{ème} colonne donne le nombre de souches reclassifiées.

1.2.1.3. Description de la procédure algorithmique de la classification bioinformatique

Un algorithme informatique permet la classification d'un pattern spoligotypique donné. Cet algorithme admet que l'évolution du locus DR se fait par des pertes successives d'espaceurs uniques ou de bloc d'espaceurs. Ainsi un pattern classifié, parmi tous les espaceurs qui lui manque, présente obligatoirement un ensemble d'espaceurs perdus qui définissent la lignée à laquelle il a été attribué. L'algorithme de classification procède en 3 étapes (figure 5).

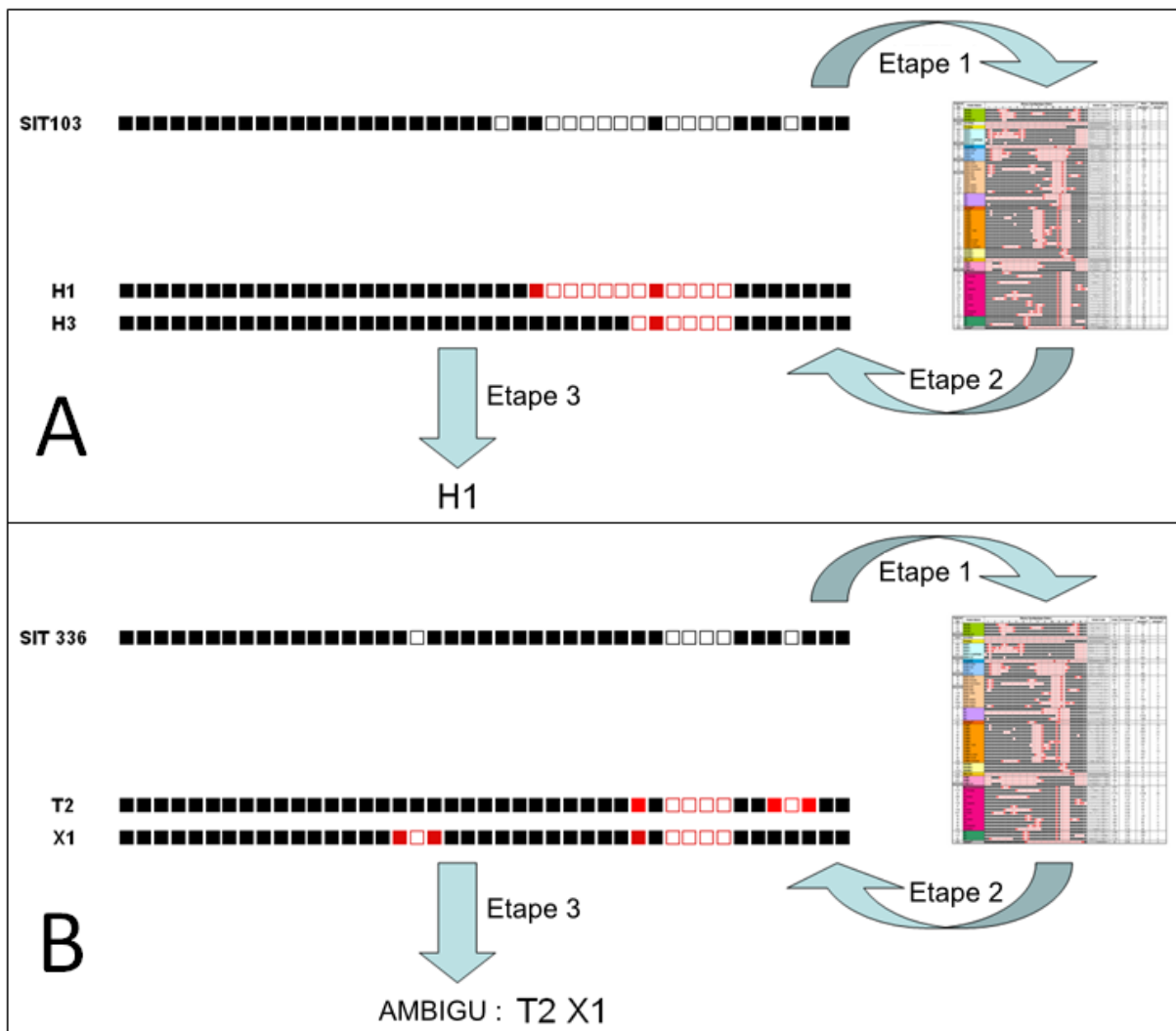


Figure 5. Procédure suivie par l'algorithme de classification

Pour commencer, lors de l'étape 1, le pattern non classifié est comparé à l'ensemble des patterns prototypes. La classification s'arrête si une seule lignée lui a été affectée ou si aucune d'elles ne semble convenir. Ensuite à l'étape 2, si plusieurs règles sont retenues, la priorité est donnée au prototype qui compte le plus

d'espaceurs délétés (figure 5B). À l'étape 3, si plusieurs prototypes sont retenus, le pattern est dit « Ambigu » (figure 5B).

1.2.1.4. Bilan de la classification des souches dans SITVITWEB

Les 67 règles ont permis la classification de $\frac{3}{4}$ des SITs, ce qui équivaut à 95% des souches spoligotypées. Environ 70% des orphelins furent également classifiés. Cependant 97 SITs sont ambigus.

1.2.2. Description de la phylogéographie de MTC à l'aide de la classification spoligotypique

Pour visualiser la phylogéographie mondiale de MTC, d'après le schéma de classification basé sur le spoligotypage, 23 sous-régions ont été considérées (figure 6).

1.2.2.1. Phylogéographie mondiale de MTC

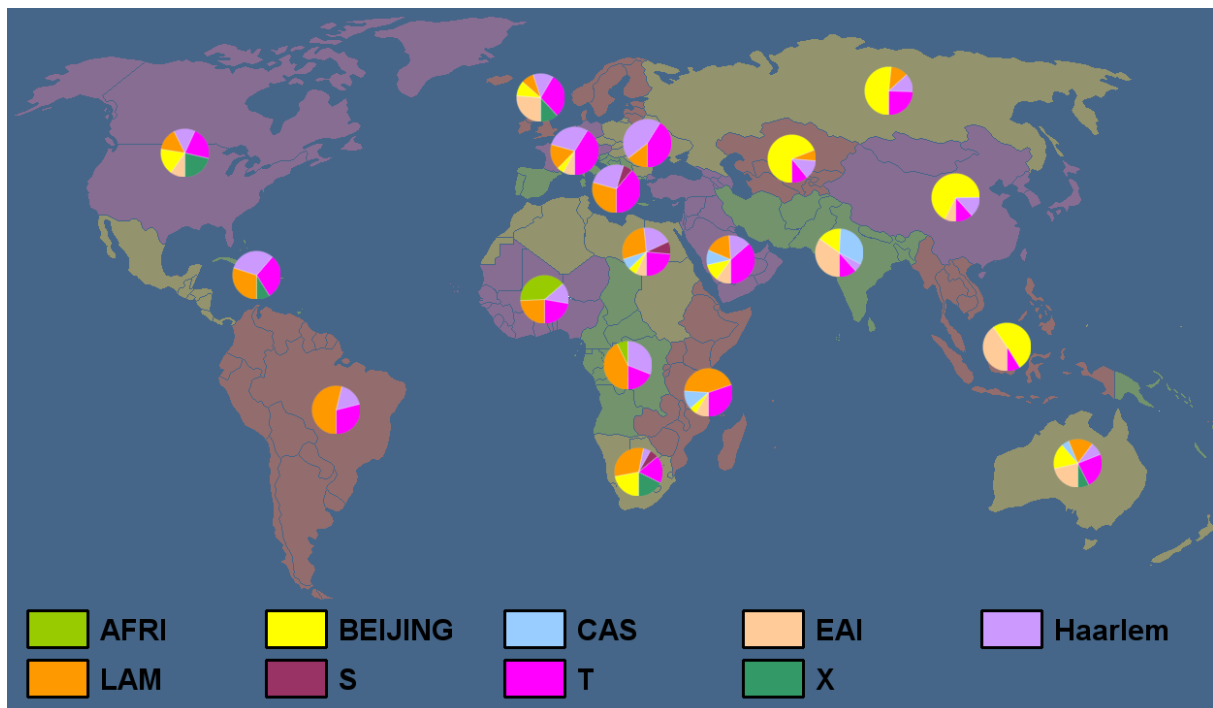


Figure 6. Distribution des clades majeurs : *M. Africanum* (AFRI), Beijing, CAS, EAI, Haarlem, LAM, S, T et X dans les diverses sous-régions du monde. Dans chaque camembert, seuls les clades représentant au moins 4% des isolats cliniques sont représentés.

La carte de la figure 7 représente la distribution des lignées majeures ; *M. Africanum* (AFRI), Beijing, CAS, EAI, Haarlem, LAM, S, T et X dans les différentes sous-régions

du monde. L'espèce *M. bovis* a été écartée car sa sur-représentativité écrasait les autres pourcentages.

Cette distribution a été faite en considérant uniquement les souches SITées. Ainsi, on voit clairement que les lignées de MTC ont une spécificité géographique. *M. africanum* se limite à l'Afrique de l'Ouest. EAI est plus largement représenté en Asie du Sud et Sud-Est. Beijing occupe préférentiellement le continent asiatique.

CAS est surtout observé en Asie du Sud. Haarlem, LAM, S, T et X recouvrent principalement, l'Europe et l'Amérique mais également l'Afrique. On observe une gradation croissante de la proportion de LAM de l'Amérique du Nord au Sud. La lignée LAM est prévalente dans les Caraïbes et en Amérique du Sud avec 29,17% et 49,34% des souches SITées. Dans les caraïbes et dans le nord de l'Amérique, les pourcentages de CAS, d'AFRI et de S étaient trop basses pour être représentés (inférieurs à 4%). T et X prédominent en Amérique du Nord (20,89% et 19,99%, respectivement).

En Europe la famille T est la plus représentée : elle englobe 35% des isolats. Elle est suivie de la famille Haarlem qui regroupe 24% des isolats SITés. T domine en Europe du Nord, du Sud, de l'Ouest alors que la lignée Haarlem est prévalente en Europe de l'Est.

L'ensemble des lignées *M. Africanum* (AFRI), Beijing, CAS, EAI, Haarlem, LAM, S, T et X, est présent sur le continent africain. L'Afrique présente alors la plus large diversité d'isolats cliniques répertoriés dans SITVITWEB. LAM s'impose dans toutes les sous-régions d'Afrique, sauf en Afrique de l'Ouest où *M. africanum* regroupe 37% des isolats cliniques SITés. *M. africanum* est également retrouvé en Afrique centrale avec 7,26% des isolats cliniques. T représente environ 20% des isolats dans toutes les sous-régions africaines. EAI et CAS excèdent 4% des isolats seulement en Afrique du Nord et de l'Est. La lignée S est retrouvée en Afrique du Nord (8,04%) et du Sud (5,81%). Beijing est également présent en Afrique ; il atteint une proportion d'environ 20% en Afrique du Sud.

La lignée Beijing s'impose en Asie. Elle représente plus de la moitié des souches dans toutes les sous-régions géographiques sauf en Asie de l'Ouest et du Sud. En Asie du Sud, EAI prédomine (33%) suivi de CAS (30,10%). EAI est également bien représenté en Asie du Sud-Est (37,62%).

1.2.2.2. Géospécificité des sous-lignées de MTC dans deux zones géographiques restreintes et distinctes

1.2.2.2.1. Description du paysage génotypique de MTC au Mozambique (Article 2)

Le Mozambique est un pays d'Afrique de l'Est de 22 millions d'habitants. C'est l'un des pays les plus touchés par la tuberculose en Afrique sub-saharienne. Au Mozambique, la tuberculose cause 27 000 décès chaque année. L'incidence et le taux de prévalence sont estimés respectivement à 431 et 504 pour 100000 habitants (WHO Report. 2009). Le Mozambique se classe 19^{ème} sur la liste des 22 pays ayant la plus forte morbidité dans le monde. Une augmentation constante du taux de la prévalence du virus de l'immunodéficience humaine (VIH) / Syndrome d'immunodéficience acquise (SIDA), rend la situation encore plus précaire. En effet jusqu'à 16,2% de la population âgée de 15 à 49 ans en 2004 est concerné par cette maladie sexuellement transmissible. Le Mozambique, avec près de 20 millions d'habitants, a des frontières géographiques avec six autres pays où la tuberculose est également endémique, c'est à dire, l'Afrique du Sud, le Swaziland, le Zimbabwe, la Zambie, le Malawi et la Tanzanie (figure 8).

Il est donc important d'étudier les géotypes prédominants de *Mycobacterium tuberculosis* qui circulent dans le pays afin de mieux comprendre l'épidémie. Cette étude a visé donc à déterminer les lignées spoligotypiques prédominantes qui causent la tuberculose au Mozambique.

1.2.2.2.1.1. Description de la collection d'isolats cliniques

L'étude a porté sur 445 isolats cliniques *M. tuberculosis*, collectés chez des patients mozambicains entre 2007 et 2008. Ces patients provenaient de diverses localités : 282 résidaient dans des régions du Sud du pays (les régions de Maputo, Gaza, Inhambane), alors que les 163 restants étaient du Nord (les régions de Nampula, Cabo Delgado et Niassa). Les patients étaient âgés de 15 à 82 ans. Plus de la moitié d'entre eux avait moins de 35 ans. L'information relative au statut HIV était disponible pour 220 patients. 22% d'entre eux fut contrôlés positifs. La majorité des séropositifs était du Sud du pays.

1.2.2.2.1.2. Description du paysage spoligotypique du Mozambique

Les 445 isolats cliniques du Mozambique ont produit 147 patterns de spoligotype43 différents. L'ensemble des résultats de génotypage a été comparé à la base de données SITVIT. 89% des souches, soient 396 isolats, présentent des patterns SITés (n=98 SITs). Dix neuf SITs contenant 28 isolats sur les 98, furent nouvellement créés, soit à l'intérieur des données de cette étude, soit en matchant avec un orphelin de la base. Parmi l'ensemble des isolats spoligotypés, 11% sont associés à des patterns orphelins (n=49). En ce qui concerne la classification spoligotypique : 96,9% des isolats furent classifiés. Les quatre lignées prédominantes sont LAM (n=165, 37%), EAI (n=132, 29,7%), T (n=52, 11,6%) et Beijing (n=31, 7%). Dans cette étude, aucun *M. africanum* et *M. bovis* n'a été observé. Il n'y a pas une différence marquée entre les souches appartenant au groupe PGG1 (41,6%) et ceux appartenant au groupe PGG2/3 (58,4%) (figure 7). Cependant, 74,6% des isolats PGG1 sont ancestraux et présentent donc de ce fait la séquence génomique TbD1+.

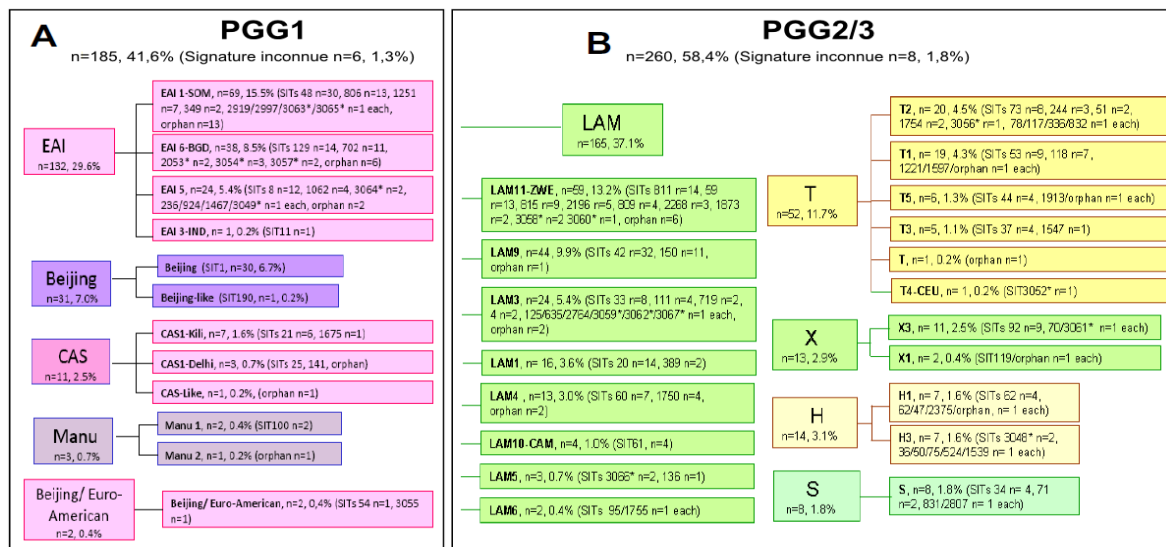


Figure 7. Description des Principaux Groupes Génétiques (PGG) de MTC au Mozambique

De même parmi les PGG2/3 on retrouve une forte prévalence de la famille LAM avec un pourcentage de 65,5%. Globalement, LAM et EAI sont les deux lignées majeures au Mozambique avec respectivement 37,1% et 29,6% des isolats

cliniques. Les autres lignées se classent selon l'ordre suivant : T (11,7%), Beijing (7%), Haarlem (3,1%), X (2,9%), CAS (2,5%), S (1,8%), Manu (0,7%). Un grand nombre des SITs prédominants appartiennent aux lignées LAM (SIT33/LAM3, SIT59/LAM11-ZWE, SIT811/LAM11-ZWE, SIT815/LAM11-ZWE) et EAI (SIT8/EAI5, SIT48/EAI1-SOM, SIT129/EAI6-BGD, SIT702/EAI6-BGD1, SIT806/EAI1-SOM) (tableau 1).

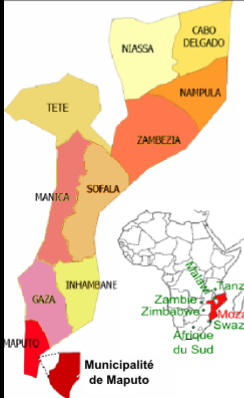
SIT (Clade)	Nombre (%) dans cette étude	% dans cette étude par rapport à SITVIT	Distribution des isolats cliniques présentant ce SIT (sous-région %)	Distribution des isolats cliniques présentant ce SIT (pays %)
1 (Beijing)	30 (6.74)	0.46	AMER-N 30.72, ASIA-SE 13.92, AFRI-S 11.76, ASIA-E 11.21, ASIA-N 8.36	USA 30.65, ZAF 11.77, RUS 8.36, JPN 8.19, VNM 5.96
8 (EAI5)	12 (2.70)	10.26	AFRI-E 26.50, EURO-N 24.79, AMER-N 24.79, ASIA-W 6.84, AFRI-S 5.13	USA 24.79, DNK 13.68, MOZ 10.26, TZA 9.40, GBR 8.55, ZMB 6.84, SAU 5.13, ZAF 5.13
20 (LAM1)	14 (3.15)	2.02	AMER-S 24.68, AMER-N 24.68, AFRI-S 12.84, EURO-S 11.40, EURO-W 8.23, CARI 6.20, AFRI-E 5.05	USA 22.94, BRA 14.29, NAM 8.95, PRT 7.07, VEN 6.06
33 (LAM3)	8 (1.80)	0.83	AFRI-S 32.60, AMER-S 23.33, AMER-N 16.77, EURO-S 14.37, EURO-W 5.73	ZAF 32.60, USA 16.56, BRA 9.48, ESP 9.27, ARG 5.94, PER 5.83
42 (LAM9)	32 (7.19)	1.26	AMER-S 30.62, AMER-N 16.71, EURO-S 13.12, EURO-W 7.21, AFRI-N 5.20	USA 15.65, BRA 10.60, COL 8.08, ITA 6.90
48 (EAI1-SOM)	30 (6.74)	7.89	EURO-N 26.32, ASIA-S 21.32, EURO-W 15.00, AFRI-E 10.00, AFRI-S 9.47, ASIA-SE 5.00	DNK 15.53, BGD 14.21, NLD 12.37, ZAF 9.47, MOZ 8.95, IND 6.05, GBR 5.26
53 (T1)	9 (2.02)	0.19	AMER-N 19.91, AMER-S 14.64, EURO-W 12.97, EURO-S 10.14, ASIA-W 8.79, AFRI-S 6.03	USA 17.54, ZAF 5.89, ITA 5.19
59 (LAM11-ZWE)	13 (2.92)	3.39	AFRI-E 67.89, AFRI-S 19.06	ZMB 27.68, ZWE 20.10, ZAF 19.06, TZA 8.36
73 (T2)	8 (1.80)	4.15	AMER-N 21.24, EURO-S 19.69, AFRI-S 13.47, EURO-W 12.44, AMER-S 10.36, AFRI-E 7.25	USA 18.65, ITA 17.62, ZAF 13.47, MOZ 5.18
92 (X3)	9 (2.02)	2.34	AFRI-S 49.09, AMER-N 24.42, AMER-S 9.61, EURO-N 5.19	ZAF 49.09, USA 21.82, BRA 5.71
129 (EAI6-BGD1)	14 (3.15)	35.90	AFRI-E 58.97, AMER-S 12.82, AMER-N 12.82, EURO-W 5.13, AFRI-N 5.13	MOZ 38.46, USA 12.82, GUF 10.26, MWI 10.26, TUN 5.13
150 (LAM9)	11 (2.47)	12.36	EURO-W 33.71, AMER-S 23.60, EURO-S 17.98, AFRI-E 13.48	BEL 24.72, MOZ 12.36, PRT 10.11, FXX 8.99, BRA 8.99, ITA 6.74, ARG 6.74, VEN 5.62
702 (EAI6-BGD1)	11 (2.47)	34.38	AFRI-E 71.88, AMER-S 15.62, CARI 6.25	MOZ 34.38, MWI 28.12, BRA 12.50, ZMB 9.38, CUB 6.25
806 (EAI1-SOM)	13 (2.92)	26.53	AFRI-S 44.90, AFRI-E 34.69, AMER-N 16.33	ZAF 44.90, MOZ 30.61, USA 16.33
811 (LAM11-ZWE)	14 (3.15)	26.92	AFRI-E 51.92, AFRI-S 38.46, AMER-N 9.62	ZAF 38.46, MOZ 28.85, ZWE 15.38, USA 9.62
815 (LAM11-ZWE)	9 (2.02)	7.83	AFRI-E 73.91, AFRI-S 21.74	ZMB 54.78, ZAF 21.74, ZWE 7.83, MOZ 7.83

Tableau 1. Description des SITs prédominants (englobant plus de 7 souches) dans cette étude, et leur distribution géographique. Pour la distribution géographique par sous-région et par pays, seules les zones avec un pourcentage de 5% et plus furent listées.

Ces SITs sont très fréquemment observés en Afrique de l'Est et du Sud, surtout dans les pays voisins du Mozambique : Zimbabwe, Zambie, Afrique du Sud, Malawi et de façon moindre en Tanzanie, Namibie et Somalie (Eldholm et al., 2006). EAI1-SOM et

LAM11-ZWE qui présentent une spécificité pour l’Afrique de l’Est, sont les deux sous-lignées spoligotypiques les plus vérifiées dans cette étude (15,5% et 13,2%). D’ailleurs, on retrouve les deux prototypes spoligotypiques de ces lignées parmi les 8 SITs les plus observés au Mozambique.

La distribution géographique des lignées génotypiques entre le Nord et le Sud du Mozambique (figure 8), montre que les lignées LAM, EAI et T sont présentes de façon assez équivalente entre le Nord et le Sud. Cependant les isolats Beijing sont plus souvent vérifiés dans le Sud (9,6%), par opposition au Nord (2,5%).



Lignée	Nord du Mozambique			Sud du Mozambique				Total
	Nampula	Cabo Delgado	Niassa	Municipalité de Maputo	Maputo	Gaza	Inhambane	
LAM	27	42	13	28	22	17	16	165
EAI	25	24	4	23	25	11	20	132
T	7	3	4	10	19	3	6	52
Beijing	3	1	0	14	7	5	1	31
H	1	2	0	4	3	2	2	14
X	0	0	0	4	4	4	1	13
CAS	1	2	1	2	3	2	0	11
S	0	0	0	2	2	3	1	8
MANU	0	0	0	0	1	0	2	3
Infection polyclonale : Beijing/Euro American Unknown	0	0	0	0	1	0	1	2
	1	1	1	5	5	0	1	14
Total	65	75	23	92	92	47	51	445

Figure 8. Distribution géographique des isolats M. tuberculosis dans 7 régions du Mozambique

1.2.2.2.1.3. Mise en évidence de deux cas d’infection polyclonale parmi les 5 isolats cliniques Manu2 de l’étude

La famille phylogénétique Manu, très rarement observée, fut initialement décrite en 2004 au cours d’une étude réalisée en Inde (Singh et al., 2004). D’autres représentants de cette lignée ont été ensuite observés à Madagascar (Ferdinand et al., 2005), en Arabie Saoudite (Al-Hajoj et al., 2007), en Tunisie (Namouchi et al., 2008), en Egypte (Helal et al., 2009). Dans l’étude réalisée en Egypte, l’occurrence de Manu atteint un plafond, avec 27% des isolats. Un nouveau SIT, SIT523 (défini par la présence des 43 espaceurs) a fait d’ailleurs son entrée dans la classification sous le nom de « Manu-ancestor » (Helal et al., 2009).

Dans l’étude du Mozambique, 5 isolats furent classifiés dans le groupe phylogénétique de Manu : 2 Manu1 (2 isolats ayant le SIT100) et 3 Manu2 (1 isolat

SIT54, 1 SIT3055, avec un orphelin). Le SIT3055 qui se caractérise par l'absence des espaceurs 4 à 12, 18, 33 à 34 semblait très singulier. De plus, en combinant le SIT1 (Beijing, absence des espaceurs 1 à 34) avec le SIT92 (X3, absence des espaceurs 4-12, 18, 33-36), deux patterns faisant partie des 10 patterns prédominants au Mozambique, on retombe sur le pattern SIT3055. Pour vérifier si les 5 Manu résultaient des cas d'infection polyclonale impliquant un isolat Beijing avec un isolat moderne PGG2/3, ils furent soumis à un typage LSP pour détecter l'absence éventuelle de la région RD105, caractéristique des souches Beijing. L'isolat ayant le pattern SIT54, présentait 2 bandes : un produit PCR intact (caractéristique des souches non Beijing) et un produit délété (caractéristique des souches Beijing). L'isolat ayant le SIT3055, ne présentait pas de bande RD105. Les patterns Manu2 SIT54 et SIT3055 sont donc le résultat de polyinfections impliquant des souches Beijing. Les 3 autres isolats Manu présentaient la région génomique RD105.

1.2.2.2.2. Description du paysage génotypique de MTC de la République de Guinée-Bissau (Article 3)

La Guinée-Bissau est un pays d'Afrique de l'Ouest qui compte 1,5 millions d'habitants. L'incidence de la tuberculose dans ce pays est moins élevée qu'au Mozambique, mais elle reste très élevée avec 233 cas pour 100 000 habitants en 2010 (WHO report, 2011). Ce pays africain est situé sur la côte Ouest, donc à l'opposé de la côte où se trouve le Mozambique. Il est donc intéressant d'étudier les clones de MTC qui circulent en Guinée-Bissau afin de pouvoir comparer la population de MTC de ce pays à celle du Mozambique.

1.2.2.2.2.1. Description de la collection d'isolats cliniques

Les différents isolats cliniques qui constituent l'échantillon de cette étude, furent isolés à l'hôpital Raoul Follereau, (hôpital de référence sur la tuberculose) chez des patients provenant de différentes régions du pays. Ainsi 414 isolats MTC ont été collectés entre 1989 et 2008. Ceux-ci ont été soumis au spoligotypage pour déterminer les lignées d'appartenance des souches étudiées. D'autres techniques de génotypage furent également utilisées ; IS6110-RFLP, le typage LSP, et la détection de l'état d'un SNP. L'application de ces techniques visait à caractériser avec davantage de précision les lignées retrouvées dans le pays.

1.2.2.2.2. Description du paysage spoligotypique de Guinée-Bissau

Les 414 isolats cliniques ont produit 120 patterns distincts dont 70 SITs et 50 orphelins.

Lignée	Sous-lignée	SIT	Lignée	Sous-lignée	SIT		
AFRI n= 195 (47.10%)	AFRI_1 n=194	<i>SIT187</i> n=96			SIT3129* n= 2		
		SIT187 n=46			SIT2440 n= 1		
		orphelin n= 26		T5 n= 3	orphelin n= 2		
		SIT536 n=6			<i>SIT44</i> n= 1		
		SIT188 n=3		T2 n= 1	orphelin n= 1		
		SIT318 n=2		Haarlem n= 37 (8.94%)	H1 n= 22	<i>SIT47</i> n= 18	
		SIT326 n=2				SIT62 n= 1	
		SIT525 n=2				SIT531 n= 1	
		SIT530 n=2				SIT2030* n= 1	
		SIT532 n=2				orphelin n= 1	
		SIT3104 n=2			H3 n= 15	<i>SIT50</i> n= 12	
		SIT3118* n= 2				SIT533 n= 2	
		SIT324 n=1				<i>SIT75</i> n= 1	
		SIT537 n=1		EAI n= 25 (6.04%)	EAI5 n= 20	SIT527 n=9	
		SIT3130 n= 1				SIT528 n= 4	
			AFRI_3 n= 1	orphelin n= 1		orphelin n= 4	
		LAM n= 75 (18.12%)	LAM9 n= 43	<i>SIT42</i> n= 36			SIT458 n= 2
SIT2201 n= 2					<i>SIT236</i> n= 1		
SIT3020* n= 2					EAI-SOM n= 4	SIT529 n= 3	
SIT766 n= 1						SIT538 n= 1	
SIT866 n= 1					EAI6-BGD n= 1	SIT129 n= 1	
orphelin n= 1				Signatures inconnues n=13 (3.14%)	Unk n= 13	orphelin n= 6	
	LAM1 n= 14					SIT20 n= 11	
				SIT534 n= 3		SIT3131* n= 2	
	LAM4 n= 9			<i>SIT60</i> n= 7		SIT523 n= 1	
				SIT828 n= 1		SIT1200 n= 1	
				orphelin n= 1		SIT1204 n= 1	
	LAM5 n= 3			<i>SIT93</i> n= 1	Beijing n= 7 (1.69%)	Beijing n= 7	<i>SIT71</i> n= 7
				orphelin n= 2			× n= 4 (0.97%)
	LAM2 n= 2			SIT604 n= 2		×1 n= 2	
	LAM3 n= 2			SIT4 n= 1			
				SIT3101* n= 1	Manu n= 4 (0.97%)	Manu2 n= 2	<i>SIT54</i> n= 1
	LAM10-CAM n= 2			<i>SIT61</i> n= 2			
T n= 53 (12.80%)	T1 n= 41	<i>SIT53</i> n= 10			Manu1 n= 1	SIT3132* n= 1	
		SIT244 n= 7			Manu-ancestor n= 1	<i>SIT523</i> n= 1	
		SIT334 n= 3		CAS n= 2 (0.48%)	CAS1-DEL n= 2		SIT3111 n= 2
		SIT804 n= 3					
		orphelin n= 4					
				SIT3100* n= 3			
				SIT196 n= 2			
				SIT522 n= 2			
				SIT535 n= 2			
				SIT230 n= 1			
				SIT521 n= 1			
				SIT611 n= 1			
				SIT801 n= 1			
				SIT888 n= 1			
			T n= 8	SIT73 n= 5			

Tableau 2. Distribution des lignées et des patterns spoligotypiques dans cette étude. Les SITs nouvellement créés sont suivis d'un astérisque. Les SITs correspondant à un prototype spoligotypique sont en italique et en gras.

Les SITs regroupent 364 isolats. Les lignées phylogénétiques suivent cette ordre d'occurrence : AFRI (n= 195, 47,10%), LAM (n= 75, 18,12%), T (n= 53, 12,80%), Haarlem (n= 37, 8,85%), EAI (n= 25, 6,04%), Beijing (n= 7, 1,68%), X (n= 4, 0,96%), Manu (n= 4, 0,97%), CAS (n= 2, 0,48%) (tableau 2). Cependant, il a été impossible d'attribuer une famille génotypique à 12 isolats (2,87%).

Ainsi *M. africanum* représente près de 50% des cas de tuberculose. Toutes les souches AFRI, à l'exception d'un orphelin (AFRI_3), appartiennent à la sous-lignée AFRI_1. Dans cette sous-lignées, deux SITs ; SIT181 (n=96) (le prototype d'AFRI_1) et SIT187 (n=46), sont les plus représentés. Ces deux SITs ont une spécificité géographique pour l'Afrique de l'Ouest (74,62% et 74,71% respectivement). Le pattern SIT527 (EAI5), semble présenter une spécificité géographique pour la Guinée-Bissau puisque qu'on en compte 9 dans l'étude alors qu'il n'est pas retrouvé ailleurs. Les SITs SIT42 (n=36, LAM9) et SIT47 (n=18, H1) sont les deux patterns PGG2/3 les plus répandus. Cependant, les patterns PGG1 (AFRI, EAI, Beijing, Manu, CAS), regroupent 56,5% de tous les isolats ici caractérisés.

1.2.2.2.3. Caractérisation de la lignée phylogénétique majeure en Guinée-Bissau

M. africanum, avec 47,1% des isolats, est prédominant en Guinée-Bissau. Pour caractériser cette famille, les typages par IS6110-RFLP et par LSP furent appliqués à 402 isolats de l'échantillon.

Ces isolats ont produit 193 profils RFLP différents. Le typage par IS6110-RFLP a regroupé 268 souches dans 61 clusters contenant 2 à 45 isolats. Les 132 patterns RFLP restants étaient uniques (orphelins). 75,6% des isolats cliniques *M. africanum* étaient en clusters par opposition aux 57,9% des isolats non *M. africanum*. Cependant, le pourcentage de souches en clusters avec moins de 6 copies d'IS6110 parmi les *M. africanum* et les non *M. africanum* était respectivement de 64,8% et 18,2%. Ainsi, les résultats de typage par IS6110-RFPL des isolats *M. africanum* en Guinée-Bissau, se caractérisent par peu de copie d'IS6110,

La délétion RD702, caractéristique de AFRI_1/ West African 2, fut évaluée pour 192 isolats cliniques *M. africanum* (sur un total n=195). Ils ont tous donné un produit PCR d'environ 700 pb, ce qui atteste de l'absence de la région RD702. Cette absence confirme bien qu'il s'agit de la lignée AFRI_1/ West African 2.

1.2.2.2.3. Conclusion sur la géospécificité des lignées

La famille « Guinea-Bissau », un groupe phylogénétique de souches MTC, fut identifiée dans des précédentes études et décrite comme présentant une spécificité géographique pour la Guinée-Bissau (Kallenius et al., 1999), (Koivula et al., 2004). Après diverses analyses issues de comparaisons génomiques, la famille « Guinea-Bissau » fut placée sur une branche phylogénétique (portant également la lignée *M. africanum*, (Mostowy et al., 2004) positionnée entre *M. tuberculosis* et *M. bovis* (Koivula et al., 2004). Cette famille regroupe des souches proches phylogénétiquement, mais renferme une hétérogénéité de caractères phénotypiques (Kallenius et al., 1999), (Koivula et al., 2004), avec une variation dans la production de niacine, dans la résistance au pyrazinamide, dans l'activité nitrate réductase. Tous ces caractères sont d'ailleurs des critères de différenciation des souches *M. bovis* des souches *M. tuberculosis*. Ces différences dans l'expression de ces caractères biochimiques ne sont point encore reliées à aucun gène (Koivula et al., 2004). La présente étude qui s'est étendue sur 20 ans, montre que la tuberculose en Guinée-Bissau est essentiellement causée par la lignée *M. africanum*, très répandue en Afrique de l'Ouest. Cette étude prouve que la famille « Guinea-Bissau » est un membre de *M. africanum* West African 2 (Gagneux et al., 2007), puisqu'elle n'a pas la région RD702. De plus cette famille a un faible nombre de copies d'IS6110, Cette lignée présente une spécificité très élevée pour l'Ouest de l'Afrique avec pour épicode la Guinée-Bissau.

Si la lignée *M. africanum* West African 2 culmine en Guinée-Bissau à l'Ouest de l'Afrique, ce sont les sous-familles EAI1-SOM (15,5%) et LAM11-ZWE (13,2%) qui prévalent au Mozambique à l'Est de l'Afrique. D'ailleurs les lignées EAI et LAM prédominent au Mozambique. Cependant en faisant une analyse géographique plus poussée, on constate que 50,3% des souches des régions du nord du pays sont des LAM contre 32,5% de EAI. Une étude réalisée au Malawi (Glynn et al., 2010), pays frontalier au Nord du Mozambique, confirme cette prédominance des LAMs dans cette zone de l'Afrique avec 51,6% des isolats du pays. Lam11-ZWE représentait 46,4% des isolats du Malawi. Dans les régions du Sud du Mozambique, LAM et EAI occupent des places équivalentes avec 29,4% et 28% des isolats cliniques respectivement. La lignée ancestrale EAI est fortement représentée sur la côte Est africaine alors que la carte de la figure 7 ne retranscrit pas ce fait. Le Mozambique

est absent de la liste des pays représentés dans la population de souches MTC, de SITVITWEB. Ceci prouve qu'à l'intérieur d'une sous-région (établie par l'UN), les pays n'ont pas les mêmes spécificités de lignées/sous-lignées.

1.3. Les limites du spoligotypage

Si le spoligotypage occupe une place importante en épidémiologie de la tuberculose, notamment pour son rôle essentiel dans la classification de Brueday (Brueday et al., 2006), il n'est pas pour autant exempt de défauts. Les infections polyclonales passent inaperçues à l'issue du spoligotypage. Aussi, ce marqueur se révèle inefficace pour la discrimination des souches Beijing.

1.3.1. Les profils spoligotypiques et les infections polyclonales (Article 4)

1.3.1.1. Généralités sur les infections polyclonales

Les infections mixtes se caractérisent par la présence simultanée d'au moins 2 souches MTC distinctes dans un cas de tuberculose. D'abord considéré comme un évènement rare, il apparaît maintenant qu'il puisse arriver plus fréquemment, en particulier dans des régions à forte prévalence de tuberculose (Baldeviano-Vidalón et al., 2005; Das et al., 2004; Warren et al., 2004; Yeh et al., 1999). Dans la plupart des études publiées, l'IS6110-RFLP ou le typage par MIRU-VNTR sont les deux principales méthodes utilisées pour l'identification des souches individuelles impliquées dans ces cas. Dans une étude antérieure (Lazzarini et al., 2007), un ensemble d'isolats, classifiés comme T, fut à la fois positif et négatif pour le LSP RD(Rio). Le spoligotype de ces souches SIT291 se retrouve sur tous les continents. Dans cette étude le SIT291 résultait d'un mélange de souches et il est impossible d'évaluer la fréquence de ce fait concernant le SIT concerné. Afin d'évaluer la probabilité qu'un pattern de SpolDB4 soit la conséquence d'une infection polyclonale, un programme informatique a été créé.

1.3.1.2. Approche informatique pour la détection des profils spoligotypiques composites

Un programme informatique fut écrit pour calculer le nombre de fois que deux patterns distincts de SpolDB4 peuvent donner un autre pattern également retrouvé dans la base.

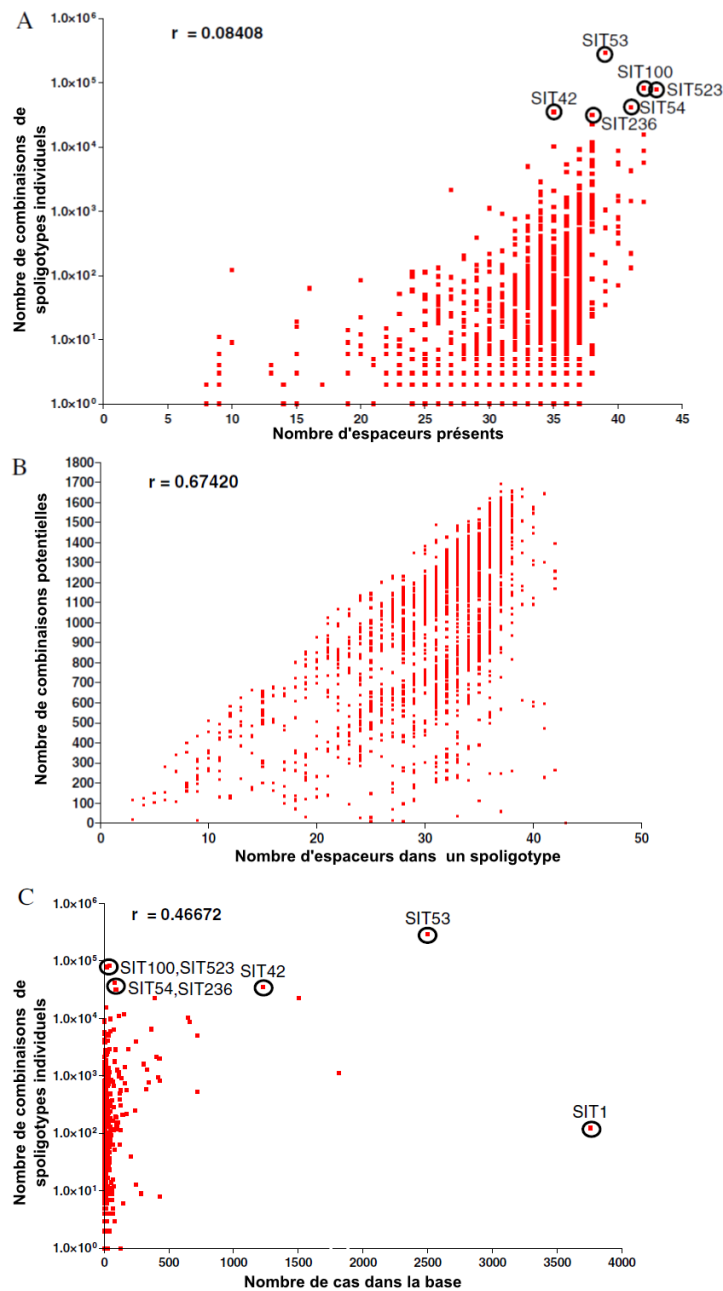


Figure 9. Approche informatique pour analyser les SITs de SpolDB4. (A) Illustration des 6 SITs ayant plus de chance de résulter du spoligotypage d'isolats extraits de prélèvements de patients touchés par une double infection tuberculeuse. Le nombre de combinaisons de spoligotypes individuels, donnant un pattern connu, est représenté en fonction du nombre d'espaceurs présents dans le spoligotype combiné résultant. (B) L'augmentation du nombre d'espaceurs présents d'un pattern entraîne une hausse de la probabilité que ce pattern soit un pattern combiné. (C) Comparaison de l'occurrence des SITs dans SpolDB4 à la chance qu'ils aient d'être les résultats de deux patterns distincts. Le nombre de combinaisons de spoligotypes individuels donnant un pattern, vérifié dans SpolDB4, est représenté en fonction du nombre de cas cliniques de tuberculose dans lesquels ce pattern est impliqué. Les cercles noirs représentent les SITs particuliers ayant fait l'objet d'une discussion.

Pour arriver à cette fin, les 1 939 SITs de SpolDB4 ont été croisés entre eux. Tous les patterns composites engendrés ont été comparés à la base pour ne retenir que ceux déjà observés, puisque ceux-ci sont susceptibles de résulter d'infections mixtes. Il en est ressorti que 1 053 SITs, sur le total des 1 939 de SpolDB4, peuvent être le produit de deux autres SITs. La probabilité qu'un spoligotype soit en réalité un composite augmente avec le nombre d'espaceurs qu'il contient. Ainsi plus le pattern de spoligotype43 présente des espaceurs, plus il est susceptible de résulter de deux souches. Cependant, cette affirmation ne doit pas être prise au pied de la lettre. Par exemple, le SIT523 (Manu-ancestor) qui a 43 espaceurs présents, peut être généré par 78 292 duos différents, alors que le SIT53 avec ses 39 espaceurs peut provenir d'un nombre plus élevé de duos 292 266. Le SIT53 se caractérise par l'absence des espaceurs 33-36. Ce groupe d'absence spécifique aux souches modernes, se retrouve chez 58% des SITs de SpolDB4. Cela explique la forte prédisposition de SIT53, (qui correspond à la lignée T), aux infections mixtes. À partir de 35 espaceurs présents, la possibilité qu'un SIT soit un composite est significative, comme le montre la figure 9A. Le nombre de combinaisons de deux patterns donnant un pattern composite présent dans la base, atteint 292 266. Cependant le nombre médian est seulement 2. Même si le plus souvent, un pattern de SpolDB4 peut être reconstitué seulement par 2 duos, l'occurrence des patterns impliqués dans ces duos joue pour beaucoup dans la probabilité de la survenance d'une infection polyclonale. La potentialité qu'à un SIT de se combiner à un autre pour former un composite est également fonction du nombre d'espaceurs (Figure 9B). Par exemple, deux SITs avec 10 et 37 espaceurs peuvent être impliqués dans un pattern SITé composite respectivement 500 et 1 692 fois.

Comme le montre la figure 9C, le SIT1 (Beijing), très couramment observé dans SpolDB4 (n=3800 isolats), a peu de chance d'être un composite, avec seulement 124 duos de SITs. Les 5 SITs qui sont les plus sujets aux infections mixtes sont SIT53 (T1) avec 292 266 possibilités, SIT100 (Manu1) avec 83 400 possibilités, SIT54 (Manu2) avec 41 688 possibilités, SIT42 (LAM9) avec 35 351 possibilités, SIT236 (EAI-5) avec 31 863 possibilités.

1.3.1.3. Description des profils spoligotypiques les plus concernés

1.3.1.3.1. Les profils les plus exposés : SIT53, SIT100, SIT54, SIT42, SIT236

Pour comprendre pourquoi les SITs SIT53 (T1), SIT100 (Manu1), SIT54 (Manu2), SIT42 (LAM9), SIT236 (EAI-5) sont les plus sujets à cacher une infection polyclonale, un arbre phylogénétique fut construit avec les SITs portés par au moins 6 souches dans SpolDB4 (n=521). L'arbre fut construit à partir de l'algorithme Camin-Sokal (proposé par le logiciel phylogénétique PHYLIP) (figure 10A). Les propriétés de cet algorithme sont décrites dans la partie « Introduction générale ». Rappelons tout de même que cette méthode obéit au principe de l'évolution unidirectionnelle du locus DR, puisqu'elle n'admet pas de réversion mutationnelle. Les 521 patterns sont donc rangés hiérarchiquement de façon parcimonieuse suivant un enchaînement de pertes d'espaceurs. Le cas des 2 SITs Manu (SIT100 et SIT54) ne seront pas discutés ici, mais dans la sous-partie qui suit. Les 3 SITs retenus (SIT53, SIT42 et SIT236) constituent dans l'arbre les nœuds basaux desquels émergent les familles T, LAM et EAI. Les patterns SIT236, SIT42, SIT53 correspondent respectivement au prototype de EAI5, de LAM9 et de T1. Ces patterns sont mis en évidence par un point noir sur l'arbre (figure 10A). Les sous-arbres ayant pour racine ces patterns sont représentés par des couleurs distinctes : un premier sous-arbre bleu enraciné en SIT236 (EAI5), un second sous-arbre vert enraciné en SIT42 (LAM9) et un troisième sous-arbre rouge enraciné en SIT53 (T).

Un algorithme fut écrit pour lister tous les patterns composites (impliquant 2 patterns différents) que peut générer un échantillon de patterns. Cet algorithme fut appliqué à 3 échantillons différents de patterns correspondant aux patterns de chaque sous-arbre. Les patterns (n=83) du sous-arbre 1 ont généré 214 patterns composites (annexe 1). 31% de ces composites sont des patterns vérifiés dans l'arbre (et 100% de ceci se retrouvent dans le sous-arbre1) (figure 10A). Le pattern composite prédominant est le SIT236 (EAI5) qui peut être généré par 1 302 duos (38,2% de l'ensemble des combinaisons). Il est suivi du SIT26 (prototype de CAS1-Delhi) qui compte seulement 4,4% des paires de patterns.

Les patterns (n=96) du sous-arbre 2 ont généré 200 patterns composites (annexe 2). 45% de ces composites sont des patterns vérifiés dans l'arbre (et 96% de ceci se retrouvent dans le sous-arbre 2) (figure 10A). Le pattern composite prédominant est le SIT42 (LAM9) qui peut être généré par 2 634 duos (57,8% de l'ensemble des

combinaisons). Il est suivi du SIT93 (prototype de LAM5) qui compte seulement 5,1% des paires de patterns.

Les patterns (n=264) du sous-arbre 3 ont généré 613 patterns composites (annexe 3). 35% de ces composites sont des patterns vérifiés dans l'arbre (et 96% de ceci se retrouvent dans le sous-arbre3) (figure 10A). Le pattern composite prédominant est le SIT53 (T1) qui peut être généré par 22 638 duos (65,2% de l'ensemble des combinaisons). Il est suivi du SIT50 (prototype de H3) qui compte seulement 3,7% des paires de patterns.

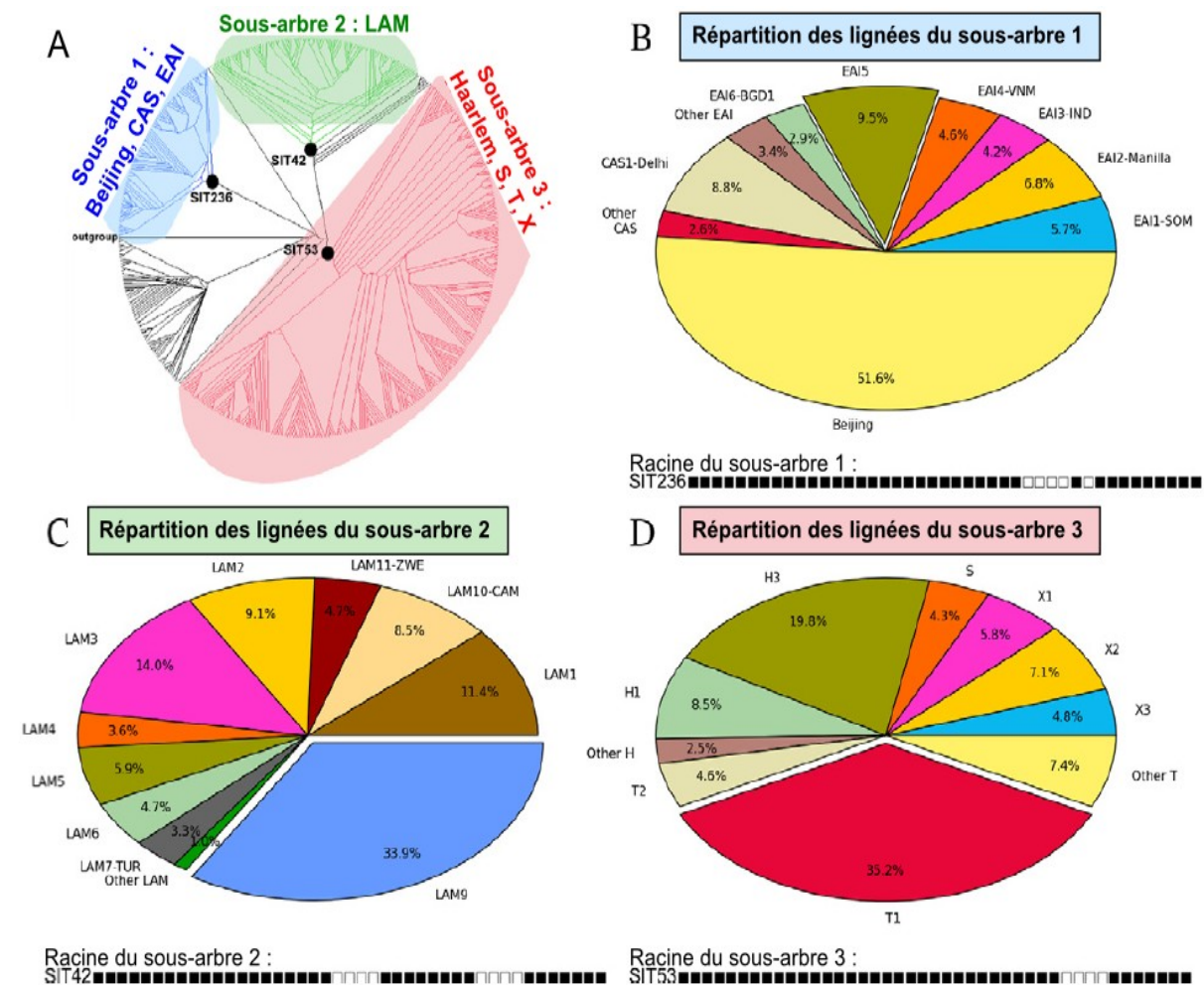


Figure 10. Analyse phylogénétique des SITs et distribution des isolats de SpoIDB4 issus de 3 échantillons distincts. (A) Analyse phylogénétique de 521 patterns (présents au moins 6 fois dans SpoIDB4), classés hiérarchiquement à l'aide de l'algorithme Camin-Sokal. L'arbre est subdivisé en 3 sous-arbres ayant chacun pour racine (point noir) un des trois SITs : SIT236 (prototype de EAI5), SIT42 (prototype de LAM9), SIT53 (prototype de T1). Le sous-arbre 1 (bleu) a pour racine le SIT236, le sous-arbre 2 (vert) a pour racine le SIT42 et le sous-arbre 3 (rouge) a pour racine le SIT53. (B) Distribution des lignées présentes dans le sous-arbre 1. (C) Distribution des lignées présentes dans le sous-arbre 2. (D) Distribution des lignées présentes dans le sous-arbre 3. Les diagrammes circulaires B, C et D furent construits avec la librairie graphique Matplotlib (<http://matplotlib.sourceforge.net>).

Ces résultats montrent clairement que la combinaison de patterns distaux dans chaque sous-arbre renvoie souvent au pattern placé à la racine de chaque sous-arbre. Les 3 SITs, SIT236 (EAI5), SIT42 (LAM9), SIT53 (T1), sont donc les patterns les plus susceptibles d'être des profils composites. Les diagrammes circulaires (figures 10B-D) donnent la distribution des lignées dans 3 échantillons d'isolats de SpolDB4. Chaque échantillon comprend les isolats de SpolDB4 décrits par les patterns présents dans un sous-arbre donné. Dans les diagrammes circulaires C (pour le sous-arbre 2) et D (pour le sous-arbre 3), LAM9 et T1 sont prédominants avec 33,9% et 35,2% des isolats. Dans le diagramme A, EAI5 est la seconde lignée présentant le plus d'isolats avec 9,5% des isolats ayant un pattern du sous-arbre 1 (= échantillon 1). En effet, par sa sur-représentation, c'est Beijing qui l'emporte avec 51,6% des isolats cliniques de l'échantillon 1. Ainsi les 3 SITs, SIT236 (EAI5), SIT42 (LAM9), SIT53 (T1), sont non seulement les patterns les plus susceptibles d'être des profils composites, mais aussi, les 3 familles EAI, LAM, T dont ils sont l'ancêtre basal, sont très bien représentées dans la base SpolDB4. Sachant que, de façon générale, le mélange de deux patterns, dévoile le pattern du plus proche ancêtre qu'ils ont en commun, et que ces 3 patterns occupent des positions basales, on peut supposer que dans la masse des isolats de SpolDB4 qui présente les trois patterns SIT236 (EAI5), SIT42 (LAM9), SIT53 (T1), il peut y avoir des cas d'isolats mixtes.

1.3.1.3.2. Le groupe phylogénétique Manu : est-ce un artéfact ?

Parmi les 5 SITs les plus concernés par les infections polyclonales, 2 appartiennent à la famille Manu (SIT54, manu2 et SIT100, Manu1). Dans l'étude menée au Mozambique, 2 isolats Manu2 (SIT54 et SIT3055) étaient le résultat de polyinfections impliquant des souches Beijing. Nous avons donc choisi d'analyser le SIT100 puisque le SIT54 a déjà été mis en cause dans une infection mixte. Les isolats avec le SIT100 sont rares dans la base SpolDB4. Ils sont répertoriés dans 4 sous-régions dans cette base. Selon SpolDB4, en Asie du Sud, en Asie du Sud-Est, en Europe du Nord, en Amérique du Nord, ils représentent 189, 14, 13 et 6 cas respectivement sur un ensemble de 10 000 cas de tuberculose (Figure 11A). Les duos de SITs capables de donner le SIT100 (Figure 11B) impliquent toujours un pattern ancestral EAI (caractérisée par la perte des espaceurs 29-32 et 34) et un pattern moderne (absence des espaceurs 33-36). Les souches modernes sont bien représentées partout dans le monde, cependant, les EAI ont une distribution

géographique plus restreinte. On remarque grâce à la figure en annexe 4 que l'incidence du SIT100 dans une région augmente avec la proportion de souches EAI dans cette même région. L'Asie du Sud, l'Asie du Sud-Est, l'Europe du Nord et l'Amérique du Nord ont les proportions les plus élevés d'isolats EAI dans SpoIDB4. De ce fait, un investigateur, en présence d'une souche avec un pattern SIT100 dans une région à forte incidence de souches EAI, doit passer au typage par MIRU-VNTR pour une vérification. Vu la rareté des patterns Manu, et les résultats des investigations menées, l'hypothèse que cette famille soit un artéfact est soulevée.

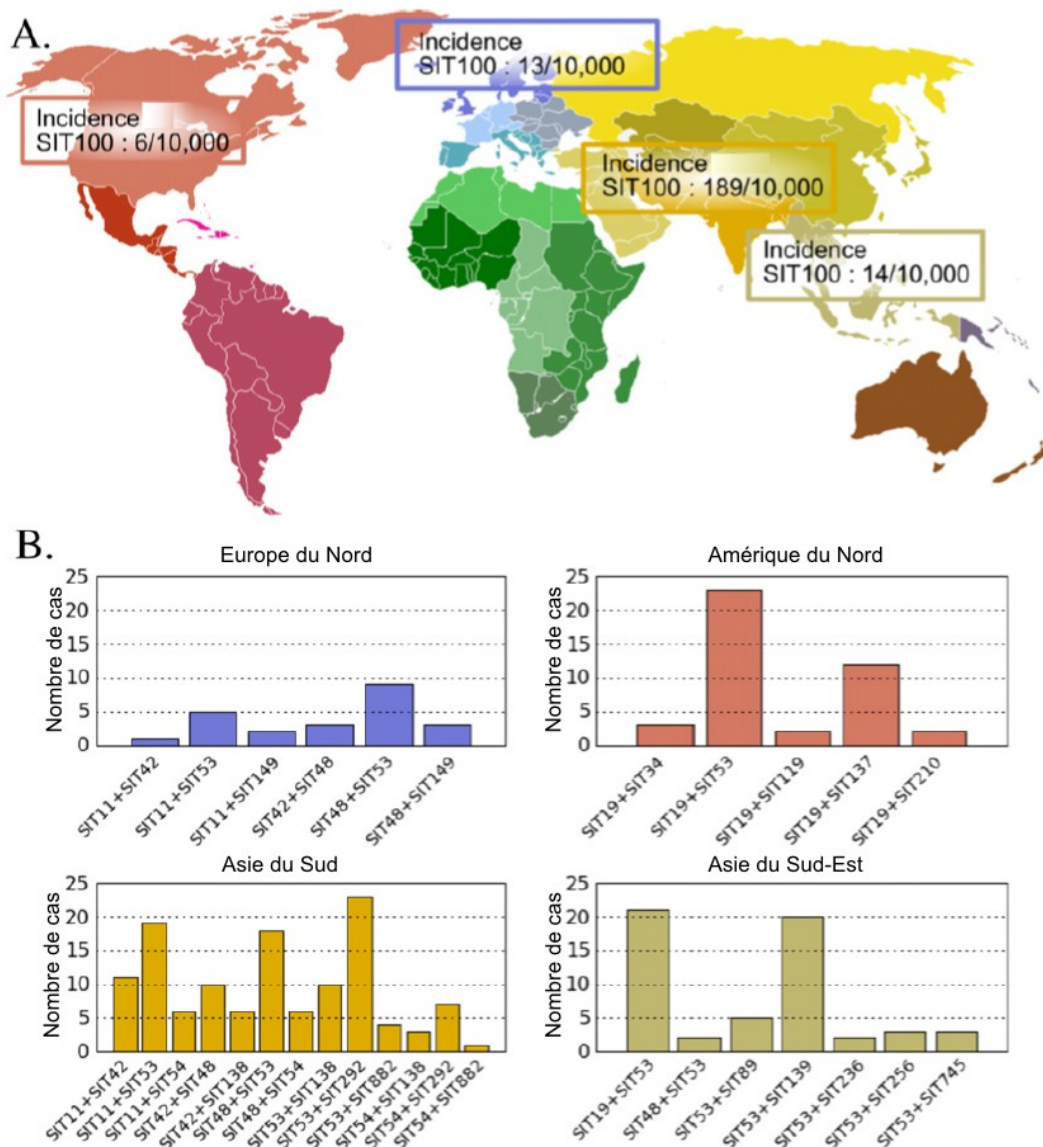


Figure 11. (A) Incidence réelle du SIT100 dans les 4 sous-régions où il est observé : Europe et Amérique du Nord, Asie du Sud et du Sud-Est. (B) Diagramme en bâtons montrant l'incidence des paires de patterns (ayant une incidence $\geq 1/10\,000$) pouvant générer le SIT100, La couleur des histogrammes fait référence à la couleur d'une sous-région dans la carte.

Si cette hypothèse ne peut pas être validée, ici, elle souligne néanmoins l'importance du typage par MIRU-VNTR pour la détection des infections tuberculeuses mixtes. Par exemple, dans une étude déjà publiée (Stavrum et al., 2009), le typage par MIRU-VNTR a révélé 54% de cas d'infection mixte (cachés par un pattern composite SIT53) dans un échantillon d'Afrique du Sud.

1.3.2. Incapacités du spoligotypage à discriminer les souches de la lignée phylogénétique Beijing (Article 5)

La lignée phylogénétique Beijing présente une spécificité pour l'Asie du Nord et de l'Est (van Soolingen et al., 1995). On la retrouve également en Afrique du Sud. L'insertion d'*IS6110* dans la région NTF (Mokrousov et al., 2005) a permis de diviser ce groupe phylogénétique en 3 sous-groupes. Le premier ne présente aucun *IS6110* dans la région NTF. Ce groupe renferme une sous-lignée ancienne qui a été d'abord nommée la « famille N ». Cette famille a causé une épidémie aux Etats-Unis en 2004 (Milan et al., 2004). Le second groupe considéré comme étant un groupe de souches Beijing modernes, présente un *IS6110* dans la région NTF. Il est bien représenté parmi les souches Beijing (Bifani et al., 2002), (Hanekom et al., 2008). Le troisième groupe est un variant moderne qui présente deux *IS6110* dans la région NTF. Celui-ci a sévit dans l'état de New-York dans les années 1990 sous une forme MDR (Bifani et al., 1996).

Les LSPs ont aussi subdivisé la famille Beijing. Alors que RD105 est absent chez tous les Beijing, trois autres LSPs ; RD181, RD142, RD150 subdivisent le groupe en 4 sous-groupes (Tsolaki et al., 2005). Cette famille qui a un niveau de dissémination très élevé est une des lignées les plus fréquemment observées dans le monde. Pour décrire ses caractéristiques, il est intéressant de comprendre sa phylogénie. Cependant, cette famille qui se définit par une grande perte d'espaces (1 à 34), ne peut être discriminée par le spoligotypage. Il est donc important d'étudier le comportement des autres marqueurs tels l'*IS6110*-RFLP et les minisatellites VNTR pour voir s'ils permettent une subdivision de cette lignée comme les LSPs et l'insertion d'*IS6110* dans la région NTF.

1.3.2.1. Description de la collection d'isolats cliniques de Sichuan en Chine

Un total de 306 isolats résistants à au moins un antibiotique furent sélectionnés parmi les souches présentant au moins une résistance à l'hôpital de Chengdu. Ces isolats proviennent de patients résidents dans la province de Sichuan en Chine. Sichuan se situe dans le Sud-Ouest de la Chine (figure 12). La prévalence de la tuberculose résistante (surtout multi-résistante ou MDR) y est plus élevée par rapport à la moyenne des autres provinces de l'Est de la Chine. En 2008, un rapport officiel a révélé que dans le Sichuan, 28,3% des frottis positifs étaient des cas de tuberculose MDR, et que 74% des cas de frottis positifs, concernaient des souches résistantes à au moins un antibiotique. Cette situation pourrait s'aggraver à cause des services de santé défectueux impliqués dans la prise en charge thérapeutique des patients tuberculeux.

Les 306 isolats ont tous été spoligotypés et MIRUtypés au format 24 loci. Seuls 274 isolats ont été soumis à l'IS6110-RFLP, car la quantité d'ADN était insuffisante pour application de la technique aux autres isolats. Ainsi, il fut possible d'évaluer le pouvoir discriminant de ces marqueurs au sein de la famille Beijing.



Figure 12. Position géographique de la province de Sichuan en Chine

1.3.2.2. Subdivision des clusters de spoligotype par les loci MIRU-VNTR

Les résultats de spoligotypage⁴³ ont été préalablement publiés pour ces 306 isolats (Zhao et al., 2011). Ceux-ci ont montré que la famille Beijing représentait 69,28% des isolats. Le spoligotypage a produit 19 clusters rassemblant 264 isolats avec 2 à 194 isolats par cluster : SIT1 (Beijing) n = 194, SIT53 (T1) n = 11, SIT334 (T1) n = 9, SIT52 (T2) n = 8, SIT632 (Beijing) n = 5, SIT54 (Manu2) n = 4, SIT1096 (Manu2) n = 4, SIT50 (H3) n = 3, SIT190 (Beijing) n = 3, SIT393 (T1) n = 3, SIT1302 (T2) n = 3, SIT3233 (lignée non définie) n = 3, SIT294 (H3) n = 2, SIT2276 (lignée non définie) n = 2, SIT2979 (Beijing) n = 2, SIT3234 (lignée non définie) n = 2, SIT3235 (T) n = 2, SIT3238 (lignée non définie) n = 2, SIT3241 (Beijing) n = 2.

Le plus large cluster ayant pour signature SIT1 a produit 18 clusters par les 5 loci VNTR, 27 clusters par les 12 loci MIRU-VNTR, 7 clusters par le format 15 loci, et 3 clusters par le format 24 loci. Ainsi par rapport à MIRU12, MIRU15 et MIRU24 ont généré beaucoup de patterns orphelins. Dans cette étude, les 12 loci classiques ont un HGDI de 0,997 alors que les formats 15 et 24 loci ont tous les deux une valeur d'HGDI de 0,999. Avec les 3 formats 12, 15 et 24, on atteint une discrimination élevée des souches avec le typage MIRU-VNTR. Le typage MIRU-VNTR est donc un bon candidat pour la subdivision de la lignée Beijing en diverses sous-lignées, subdivision à laquelle l'IS6110-RFLP et les LSPs se sont déjà attelés.

1.3.2.3. Subdivision des isolats Beijing et non-Beijing par les résultats de l'IS6110-RFLP

L'échantillon de 274 isolats typés par l'IS6110-RFLP a produit 5 clusters et 262 patterns uniques. L'HGDI de cette méthode fut évalué à 0,999. Les 190 souches Beijing furent subdivisées en 2 clusters et 186 patterns orphelins. Les 84 isolats non Beijing ont donné 3 clusters et 76 patterns uniques. La taille des clusters varie entre 2 et 3 isolats. Le nombre de copies d'IS6110 varie entre 7 et 17 parmi les souches Beijing, et 1 et 16 parmi les non Beijing.

1.3.2.4. Conclusion sur la discrimination des souches Beijing

Les LSPs et le nombre d'insertions d'IS6110 dans la région NTF ont déjà prouvé qu'il y a des sous-groupes au sein de la famille Beijing. Cette étude réalisée dans le Sichuan montre aussi un très fort pouvoir discriminant de l'IS6110-RFLP et

des minisatellites MIRU-VNTR dans l'échantillon global de Sichuan. Ceci atteste de la capacité de ces marqueurs à retranscrire une variabilité à l'intérieur de cette famille, fait qui n'est pas si évident avec le spoligotypage. Les VNTR ont déjà montré leur potentiel à émettre des hypothèses intéressantes sur l'évolution, sur les transmissions des souches Beijing, et ce, grâce à des analyses phylogénétiques et des analyses comparatives basées sur des caractéristiques phénotypiques d'isolats cliniques (Wada et al., 2008). Beijing et LAM sont les deux lignées (strictement pathogènes à l'homme) les plus distribuées dans le monde (Zozio, 2011). Si la lignée LAM s'est scindée en plusieurs sous-groupes (Brudey et al., 2006), (Gibson et al., 2008) au cours de son évolution, il est très fortement probable que la lignée Beijing l'ait fait aussi. Une classification basée sur les minisatellites MIRU-VNTR serait donc profitable pour se rendre compte de ces subdivisions phylogénétiques.

1.4. Quel est le potentiel réel des loci VNTR dans l'épidémiologie et la phylogénie de la tuberculose ? (Article 1)

1.4.1. La variabilité des loci VNTR à l'intérieur des clones spoligotypiques

L'index de Hunter-Gaston (HGDI) est ici utilisé pour évaluer la variabilité de chaque locus MIRU-VNTR dans 10 lignées génotypiques majeures : *M. africanum*, Beijing, BOVIS, CAS, EAI, Haarlem, LAM, S, T, X. Les calculs ont été réalisés sur 3178 isolats cliniques de SITVITWEB présentant à la fois des données de spoligotypage et de MIRU12. Pour cette étude, les données relatives aux clades T1, T2, T3, T4, et T5 ont été exclues car il s'agit de sous-lignées mal définies. De même, les isolats ayant des patterns MRU12 avec des nombres de répétitions supérieurs à 9 ont été écartés.

Le diagramme en bâtons de la figure 13 prouve que les minisatellites MIRU-VNTR ont une variabilité différente d'une famille à une autre. Par exemple, le MIRU-10 (ou locus 10) a un HGDI global très élevé (0,73). Chez *M. africanum*, l'HGDI de ce locus atteint 0,78 alors qu'au sein de la lignée S, il chute à 0,24. Le MIRU-24 est constant chez Beijing, CAS, Haarlem, LAM, S et T (une seule copie). MIRU-2 et MIRU-23 sont constants chez les isolats appartenant à CAS, (respectivement 2 copies et 5 copies). Face à l'évidence de la variabilité spécifique des loci MIRU-VNTR au sein des lignées spoligotypiques, la perspective de trouver un lien entre patterns MIRU12 et lignées phylogénétiques de MTC, devient plus attrayante et envisageable.

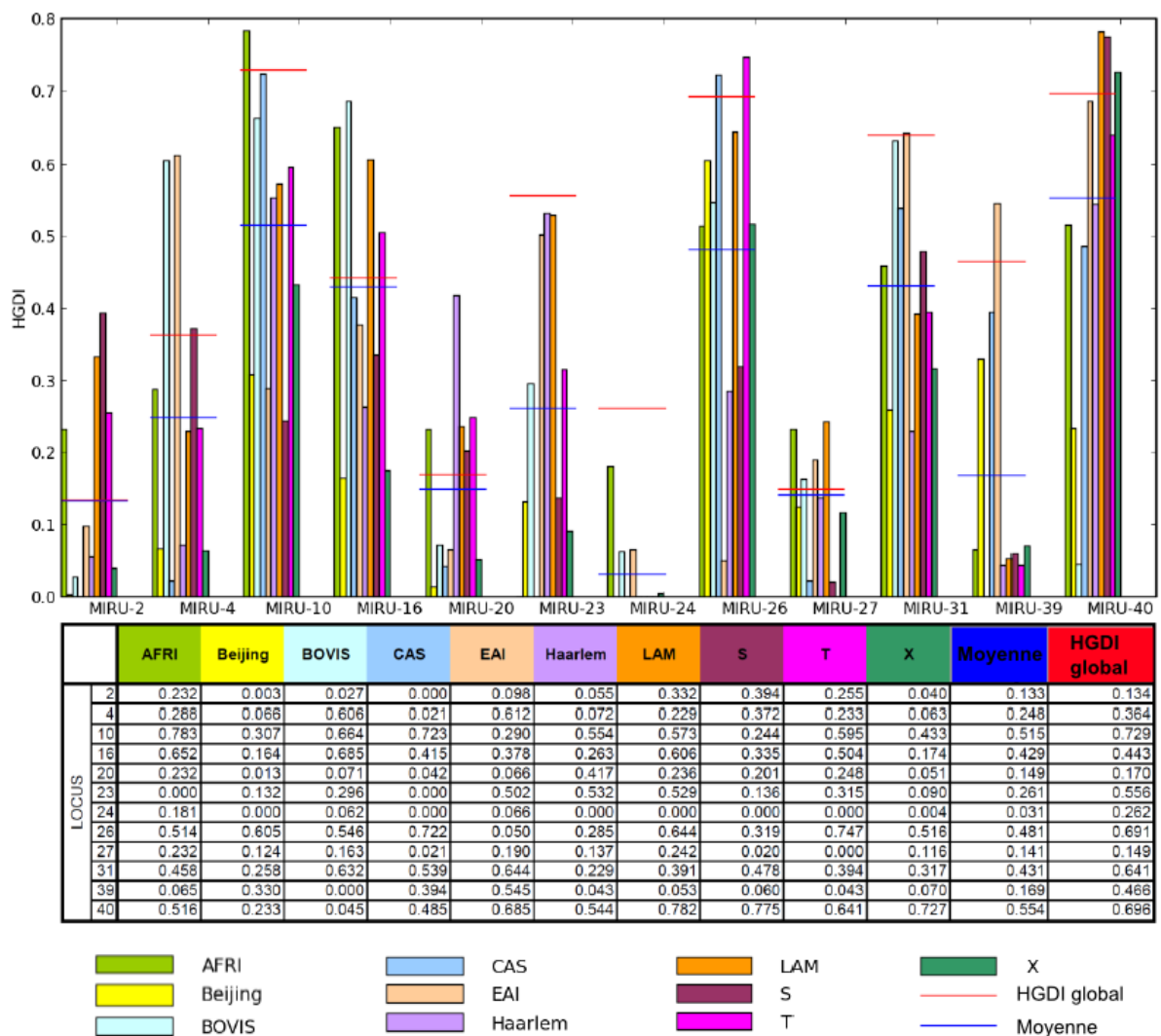


Figure 13. Diversité allélique (HGDI) des 12 minisatellites MIRU-VNTR parmi 10 lignées génotypiques (*M. africanum*, Beijing, Bovis, CAS, EAI, Haarlem, LAM, S, T, et X) du complexe *Mycobacterium tuberculosis*. Les lignes horizontales rouges correspondent à la valeur de l'HGDI global obtenue pour chaque locus. Les lignes bleues désignent les valeurs moyennes d'HGDI, retrouvées individuellement dans les 10 lignées génotypiques, pour un locus donné. Le détail des HGDI est reporté dans le tableau qui suit. Les calculs ont été réalisés sur 3 178 isolats cliniques de SITVITWEB présentant à la fois un spoligotype et un pattern MIRU12. Le diagramme en bâtons est réalisé avec la librairie graphique Matplotlib (<http://matplotlib.sourceforge.net/>).

1.4.2. Hypothèse d'une variabilité spécifique des loci MIRU-VNTR au sein de différentes lignées phylogénétiques

La figure 14 illustre la distribution des 1 018 patterns MIRU-VNTR (format 12 loci) à l'intérieur de 10 lignées spoligotypiques majeures : *M. africanum*, S, T, CAS, *M. bovis*, X, Beijing, Haarlem, EAI, LAM. Une lignée spoligotypique fut attribuée aux 3 178 isolats cliniques concernés grâce aux patterns spoligotypiques qui leur sont associés. Pour augmenter la véracité de cette distribution, les données relatives aux

sous-lignées mal définies T1, T2, T3, T4, et T5 ont été exclues du groupe des isolats T.

L'ordre classique des 12 loci MIRU-VNTR a été modifié pour mieux apprécier leur distribution. Dans ce but, les loci MIRU ont été rangés dans l'ordre croissant de l'HGDI moyen des loci dans les différentes lignées (représenté par les lignes bleues de la figure 13). Ces 1 018 patterns réorganisés, sont considérés comme des nombres, et positionnés sur l'axe des abscisses de la figure 14.

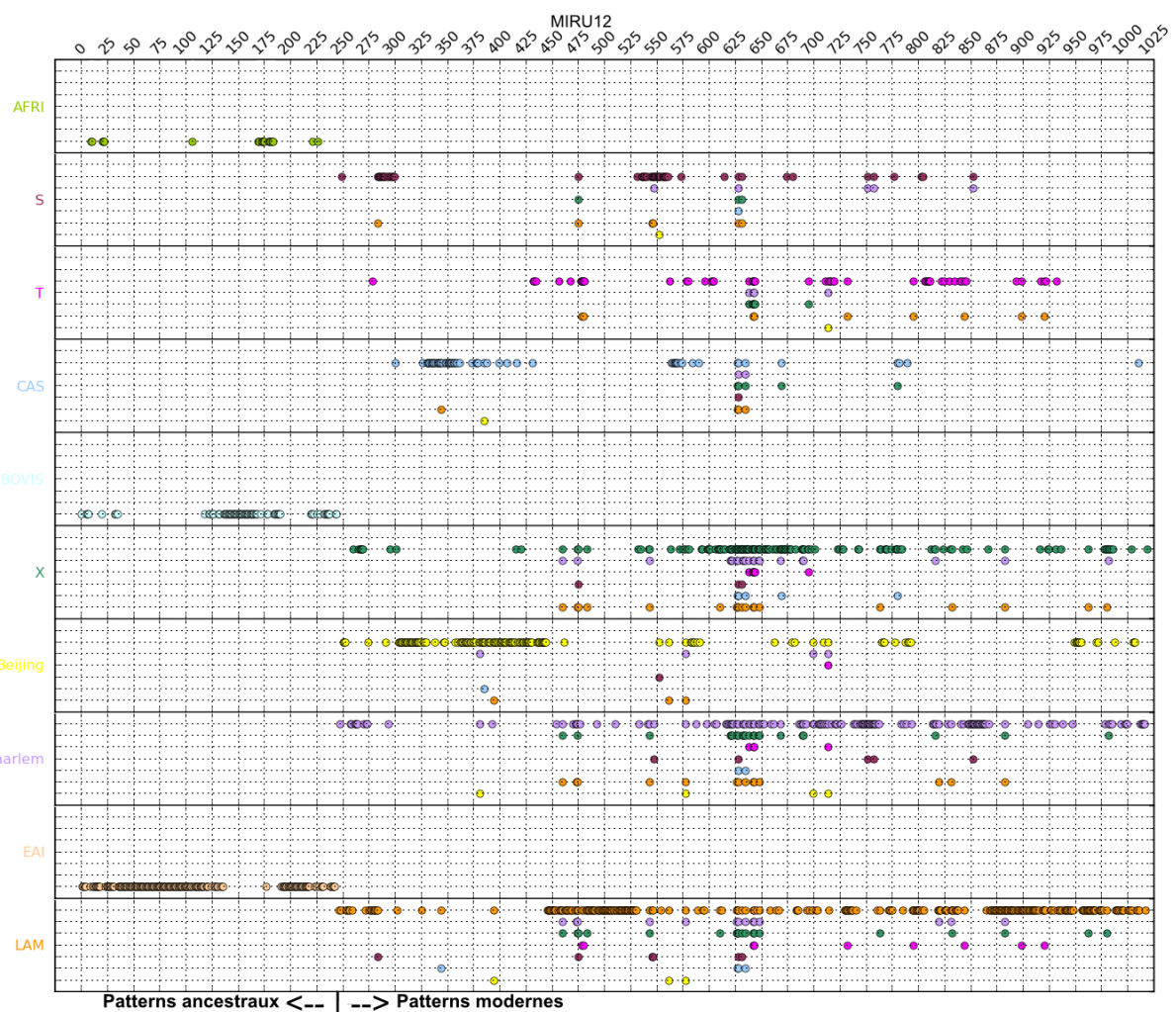


Figure 14. Distribution des patterns MIRU12 réorganisés, à l'intérieur de 10 lignées spoligotypiques majeures : *M. africanum*, S, T, CAS, *M. bovis*, X, Beijing, Haarlem, EAI, LAM.

On observe ainsi une bonne séparation des isolats ancestraux et modernes. Certains patterns se retrouvent dans plusieurs lignées spoligotypiques. Ces patterns non spécifiques se retrouvent surtout dans les familles modernes PGG2/3, LAM, Haarlem, X et T mais jamais chez *M. africanum*, *M. bovis* et EAI. Ce manque de spécificité des patterns MIRU-VNTR a 3 causes probables : (i) une évolution

convergente du marqueur MIRU-VNTR, (ii) une définition de la classification spoligotypique inexacte pour les lignées modernes PGG2/3, (iii) des résultats de génotypage (spoligotypiques ou MIRU-VNTR) erronés. Vu que ces cas de patterns MIRU12 non spécifiques se limitent seulement aux lignées modernes, les deux premières causes (i) et (ii) sont les plus probables.

Autre fait intéressant : les patterns MIRU12 de chaque lignée, tendent à se regrouper en bloc à des positions différentes de l'axe. Les 125 premiers patterns concernent principalement des EAI, puis du 130^{ième} au 175^{ième} on retrouve les *M. bovis* etc. L'axe peut ainsi être divisé en 3 grandes intervalles : (i) le sous-groupe 1 englobe les 250 premiers patterns appartenant aux isolats ancestraux (EAI, *M. bovis*, *M. africanum*), (ii) le sous-groupe 2 s'étend du 300^{ième} au 450^{ième} et inclut Beijing et CAS, (iii) le sous-groupe 3 s'étend du 525^{ième} au 1025^{ième} et comprend les lignées modernes; LAM, S, X, Haarlem et T. Dans cette subdivision se dévoile une logique. En effet, cette subdivision est superposable à celle de MTC basée sur les SNPs *katG* et *gyrA* (Sreevatsen et al., 1997) et à la région de délétion Tbd1 (Brosch et al., 2002). Le sous-groupe 1 peut être identifié au PGG1/Tbd1+, le sous-groupe 2 au PGG1/Tbd1- et le sous-groupe 3 au PGG2-3/Tbd1- qui englobe les souches de la grande lignée Euro American (Gagneux et al., 2007).

2. Utilisation de l'intelligence artificielle pour capturer les informations cachées dans la base de données SITVITWEB (Article 6)

Les deux méthodes basées sur la PCR les plus couramment utilisées pour l'identification des clones circulants sont le spoligotypage⁴³, et le typage par MIRU-VNTR. Les minisatellites MIRU ont un pouvoir discriminant plus élevé que le spoligotypage, mais ces deux marqueurs moléculaires sont souvent utilisés de façon combinée pour éviter les mauvaises interprétations que peuvent provoquer les phénomènes de convergence (Warren et al., 2002), (Abadia et al., 2010). Ces deux marqueurs ne présentent pas le même taux de mutations (Supply et al., 2000, 2011), et ne partagent pas le même mécanisme de variation génomique (Warren et al., 2002), (Thorne et al., 2007), (Billamas et al. 2007). Le spoligotypage⁴³, et le typage par MIRU-VNTR constituent les piliers de la base de données SITVITWEB (Demay et al., 2012). Quoique beaucoup utilisés, on connaît peu de choses de ces

marqueurs. Vu que les mécanismes qui se cachent derrière le polymorphisme de ces marqueurs n'ont pas encore été percés, il est difficile d'avoir, par exemple, une interprétation juste et complète des arbres inférés lors des études phylogénétiques. On devine donc que les nombreuses informations de typage stockées dans SITVITWEB sont sous-exploitées. L'utilisation des réseaux bayésiens constitue donc une bonne alternative pour extraire des informations cachées de cette grande base de données génotypiques.

2.1. VNTR-TBbayes, une application WEB pour l'aide à la classification, à la complétude des données de typage manquantes et à l'attribution d'une origine géographique aux souches de MTC

2.1.1. Intérêt des réseaux bayésiens pour MTC

Le modèle bayésien, VNTR-TBbayes, permet, en utilisant des variables adéquates, de vérifier les informations de typage, de compléter les informations manquantes, de donner une origine à un profil de génotypage (MIRU et/ou spoligotype43), d'attribuer une lignée spoligotypique à un profil de génotypage (MIRU et/ou spoligotype43). Ainsi à partir de données observées, VNTR-TBbayes attribue une probabilité aux variables inconnues. Cet outil constitue donc une aide à la décision pour la classification des souches, la vérification des résultats de typage et des informations relatives à l'origine géographique d'une souche.

2.1.2. Choix des variables

Dans la partie traitant de la pertinence et du potentiel des données recueillies dans SITVITWEB, on a vu que le spoligotypage est utile à la classification des souches MTC, que dans cette classification, les lignées spoligotypiques présentaient des spécificités phylogéographiques, et que les minisatellites MIRU-VNTR ont un potentiel de classification. Cette perspective de classification basée sur les loci MIRU-VNTR est d'autant plus attrayante que le spoligotypage semble parfois être soumis à des évolutions convergentes, masque les infections polyclonales et n'est pas discriminant pour certaines familles spoligotypiques (cas de la famille Beijing). Les variables considérées pour la construction de modèle bayésien décrivant le complexe *M. tuberculosis* sont donc :

- Une variable pour l'origine géographique de l'hôte qui prend 15 états relatifs aux 15 sous-régions géographiques considérées,

- Une variable pour chacun des espaceurs du spoligotypage, soit un total de 43 variables prenant deux états : absent ou présent,
- Une variable pour chacun des 12 loci MIRU-VNTR, soit un total de 12 variables dont les attributs varient entre 0 et 9.
- Une variable pour la lignée prenant pour état un ensemble de lignées/sous-lignées.

Le modèle établit donc des relations de causalité entre les variables, ce qui permet, entre autres, de comprendre l'impact des variations des loci MIRU-VNTR sur la classification.

2.1.3. Description des échantillons utilisés pour construire et tester le modèle

Les isolats sélectionnés pour faire l'apprentissage de structure et pour ensuite tester le modèle bayésien devaient contenir les informations de spoligotypage et de MIRU12, d'origine géographique ainsi que de classification.

Un premier échantillon, ayant permis la construction du réseau, a été extrait de SITVITWEB et compte 3 818 isolats. On y retrouve les lignées suivantes : *M. africanum* (n=20), Beijing (n=740), *M. bovis* (n=190), *M. canettii* (n=3), CAS (n=62), EAI (n=217), Haarlem (n=460), LAM (n=663), Manu (n=6), *M. microti* (n=2), *M. pinnipedii* (n=4), S (n=98), T (n=856), X (n=496), Zero copy (n=1). Ces souches sont originaires de 15 sous-régions : les Caraïbes, l'Asie de l'Est, l'Europe de l'Est, l'Afrique Centrale, l'Amérique du Nord, l'Amérique du Sud, l'Asie du Nord, l'Europe du Nord, l'Asie du Sud-Est, l'Afrique du Sud, l'Asie du Sud, l'Europe du Sud, l'Afrique de l'Est, l'Asie de l'Est et l'Europe de l'Est. Vu la grande quantité de données (n=3 818) et de variables (n=57), la sollicitation du logiciel GeNIe (Graphical Network Interface, disponible en ligne <http://genie.sis.pitt.edu/about.html>) a été nécessaire pour faire l'apprentissage de structure et donc dévoiler les relations de dépendance existant entre les variables considérées.

L'échantillon de test se compose de 880 isolats répartis dans les diverses lignées spoligotypiques comme suit : (CAS n=61), Beijing (n=80), EAI (n=101), *M. bovis* (n=15), *M. africanum* (n= 5), Manu (n=1), Haarlem (n=133), LAM (n=187), T (n=213), X (n=56) et S (n=28). Les tests permettant de juger de la pertinence de la prédiction des lignées par VNTR-TBbayes ont donc concerné les lignées : CAS, Beijing, EAI, *M. bovis*, *M. africanum*, Manu, Haarlem, LAM, T, X et S.

2.2. Evaluation de la pertinence de VNTR-TBbayes dans les prédictions

2.2.1. Evaluation de la pertinence de VNTR-TBbayes dans l'assignation des lignées

2.2.1.1. Seul le spoligotype est placé en évidence

Pour évaluer la pertinence de VNTR-TBbayes à prédire la lignée à partir du spoligotype, tous les patterns de spoligotype⁴³ de l'échantillon de test ont été soumis à VNTR-TBbayes. Ainsi, dans ce premier test, seul le spoligotype⁴³ a été considéré comme évidence. La sensibilité de la prédiction dans ce cas varie entre 82% et 100% pour toutes les lignées à l'exception des patterns CAS et Beijing qui ne sont jamais détectés, (sensibilité de 0%) (tableau 3). Dans la plupart des cas, les isolats CAS et Beijing sont prédits EAI, ce qui contribue à baisser la spécificité de l'application pour la lignée EAI (47,9%). On remarque par exemple que 4 isolats Beijing ont été prédits AFRI et un autre a été prédit Haarlem. Cela peut être une mauvaise classification de VNTR-TBbayes ou le résultat d'une convergence dans le locus DR. On sait que le locus DR est parfois victime de longues délétions. Dans le cas des souches classifiées *M. microti*, la perte des espaceurs 1 à 36 est causée par la délétion (LSP) MID1 qui serait causée par un IS6110 (Warren et al., 2002) (Brodin et al., 2002). Ce mécanisme de délétion peut être corrélé avec la délétion de 7 399 pb, chez les isolats Beijing, délétion (RD207) qui emporte avec elle une partie du locus DR (c'est-à-dire les espaceurs 1 à 34) ainsi que 7 gènes et un IS6110 (Tsolaki et al., 2005). Mauvaise prédiction de VNTR-TBbayes, convergence due aux recombinaisons entre deux IS6110, ou erreur dans les résultats de génotypage, les trois faits sont plausibles.

Lignée prédictive																																
Lignée de Brudey et al., (2006)	CAS		Beijing		EAI		M. bovis		M. africanum		Manu		Haarlem		LAM		T		X		S		Sensibilité (%)									
	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M	S	M						
Evidence	54	61	4	80	34	11	11	16	1	1	1	1	1	1	1	1	1	1	1	2	2	0	88.5	100	0	96.3	100					
CAS	54	61	4	80	34	11	11	16	1	1	1	1	1	1	1	1	1	1	1	2	2	0	88.5	100	0	96.3	100					
Beijing	77	76	77	76	76	3	3	1	1	1	1	1	1	1	1	1	1	1	1	2	2	0	96.3	100	0	96.3	100					
EAI	101	97	100	100	100	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100	96	99	100	96	99					
M. bovis	15	15	15	15	15	2	2	15	15	15	15	15	15	15	15	15	15	15	15	15	15	100	100	100	100	100	100					
M. africanum	2	5	3	5	5	2	5	3	5	5	5	5	5	5	5	5	5	5	5	5	5	100	60	100	100	60	100					
Manu	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100	0	100	100	0	100					
Haarlem	1	132	80	133	5	13	13	13	5	5	5	5	5	5	5	5	5	5	5	32	1	2	99.3	60.2	100	99.3	60.2	100				
LAM	1	8	184	144	185	3	21	2	2	2	2	2	2	2	2	2	2	2	2	7	6	6	98.4	77	98.9	98.4	77	98.9				
T	2	17	3	3	43	6	207	125	201	3	24	3	24	3	24	3	24	3	24	3	24	2	97.2	58.7	94.4	97.2	58.7	94.4				
X	13	13	13	13	13	2	2	13	13	13	13	13	13	13	13	13	13	13	13	56	30	56	100	53.6	100	100	53.6	100				
S	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	23	12	25	82.1	42.9	89.3	82.1	42.9	89.3				
Spécificité (%)	0	100	100	90.6	100	47.9	100	100	100	75	100	26.3	75	100	5.9	0	50	99.3	66.1	97.8	95.8	72.4	95.4	98.6	69.1	99	94.9	31.6	94.9	95.8	52.2	100

Tableau 3. Évaluation de l'efficacité prédictive de VNTR-TBbayes. Les prédictions sont réalisées à partir du spoligotypage (S) seul, du MIRU12 seul (M) ou des deux combinés (SM) d'un échantillon pour lequel les clades ont été considérés comme étant inconnus (n=880). La sensibilité et la spécificité ont été calculées pour chaque lignée.

* La spécificité évalue l'efficacité du modèle dans sa capacité à reconnaître les vrais négatifs, (donc ne pas prédire les lignées non attendues).

**La sensibilité évalue l'efficacité du modèle dans sa capacité à reconnaître les vrais positifs, (donc à prédire la lignée attendue).

2.2.1.2. Seul le MIRU12 est placé en évidence

Lorsque les MIRUs sont utilisés seuls comme évidence, la sensibilité est satisfaisante pour les lignées PGG1 *M. bovis* (100%), EAI (96%), Beijing (96.3%), CAS (88.5%), à l'exception de *M. africanum* (60%), et Manu (0%); Cependant, vu qu'il y avait un seul isolat Manu dans la base de test, toute interprétation doit être évitée. Ainsi, l'application a une bonne sensibilité pour les lignées PGG1 (88,5% à 100%) à l'exception de *M. africanum* (60%). Contrairement aux spoligotypes seuls où les lignées modernes PGG1 CAS et Beijing ne sont pas détectées, les loci MIRU-VNTR seuls arrivent à proposer ces lignées lorsqu'elles sont attendues. Cependant les MIRUs seuls, ne permettent pas toujours la détection des isolats PGG2/3, pour lesquels la sensibilité est comprise entre 42,9% et 77%. La sensibilité pour ces lignées, avec les patterns de spoligotypage seuls était comprise entre 82,1% et 100% (tableau 4). Cependant si l'on considère le schéma de classification basé sur les LSPs (Gagneux et al., 2007), la sensibilité atteint 100% pour le groupe Euro American qui regroupe les familles spoligotypiques Haarlem, LAM, T, X et S ayant en commun la perte des espaceurs.

2.2.1.3. Le spoligotype et le MIRU12 sont placés en évidence

En considérant les valeurs de spoligotypage et de MIRU12, la sensibilité de VNTR-TBbayes pour chaque lignée spoligotypique augmente. Elle atteint 100% pour CAS, Beijing, *M. bovis*, *M. africanum*, Manu, Haarlem, X; 99% pour EAI et LAM; 94.4% pour T et 89.3% pour S (tableau 3).

2.2.1.4. Conclusion sur la pertinence de VNTR-TBbayes dans l'assignation des lignées à partie des données issues du spoligotypage et/ou du typage des minisatellites MIRU-VNTR

Pour atteindre une bonne sensibilité dans la prédiction des lignées, les deux marqueurs spoligotypage⁴³ et MIRU12 sont complémentaires. Les lignées PGG1 Beijing et CAS ne sont pas détectées par les spoligotypes seuls alors qu'elles le sont avec MIRU12. Cela prouve que VNTR-TBbayes s'appuie davantage sur la valeur des loci MIRU pour prédire les lignées. Cependant, chez les lignées PGG2/3 la prédiction est meilleure avec les spoligotypes qu'avec les MIRU12 seuls. Ceci s'expliquerait par le fait que les loci MIRUs n'arrivent pas à s'approprier la classification spoligotypique au sein des PGG2/3. À partir des minisatellites, VNTR-

TBbayes fait beaucoup de confusion entre les souches Haarlem, LAM, T, X et S, ce qui met un doute sur les règles spoligotypiques qui définissent ces familles. La partie 3 de ce chapitre qui traite d'une classification basée sur les données de typage MIRU-VNTR apporte des explications à ce fait.

2.2.2. Evaluation de la pertinence de VNTR-TBbayes dans l'assignation des sous-lignées connaissant la lignée

Vu que l'on suspecte des incohérences dans le schéma de classification de Brudey (Brudey et al., 2006) (Demay et al., 2012) pour les lignées modernes PG2/3, seuls les isolats appartenant au groupe PGG1 ont été soumis au test pour la prédiction des sous-lignées connaissant le spoligotype, le MIRU12 et la lignée. Les résultats sont rapportés dans le tableau 4. Comme le montre ce dernier, VNTR-TBbayes s'avère très fiable pour prédire la sous-famille quand la famille est connue. La sensibilité est de 100% pour 11 des 14 sous-lignées testées. Cependant la sensibilité est faible pour EAI5 (73.9%), et nulle pour EAI6-BGD1. L'explication est simple : l'échantillon ayant servi à l'apprentissage de structure de VNTR-TBbayes ne contenait pas d'isolats EAI6-BGD1.

Sous-lignée de Brudey et al., (2006)	Sous-lignée prédite														Sensibilité (%) ***	
	AFRI1*	BOV1	BOV2	CAS1-Delhi	CAS1-Kili	CAS2	EAI1-SOM	EAI2-Manila	EAI2-nonthaburi	EAI3-IND	EAI4-VNM	EAI5	EAI6-BGD1	EAI8-MDG		
AFRI1*	3															100
BOV1		11														100
BOV2			2													100
CAS1-Delhi				39												100
CAS1-Kili					7											100
CAS2						4										100
EAI1-SOM							12									100
EAI2-Manila								24								100
EAI2-nonthaburi								1	4							80
EAI3-IND										10						100
EAI4-VNM											1					100
EAI5											6	17				73.9
EAI6-BGD1													11			0
EAI8-MDG															8	100
Spécificité (%)**	100	100	100	100	100	100	100	96	100	100	14.3	60.7	0	100		

Tableau 4. Evaluation de la pertinence de la prédiction des sous-lignées connaissant les lignées, le spoligotype et le MIRU12. Ce test est réalisé sur un échantillon de 160 isolats cliniques PGG1 provenant de SITVIT. La spécificité et la sensibilité de VNTR-TBbayes sont calculées pour chaque sous-lignées.

* AFRI1 (Brudey et al., 2006) correspond à *M. africanum* West African 2 (Gagneux et al., 2007).

** La spécificité évalue l'efficacité du modèle dans sa capacité à reconnaître les vrais négatifs, (donc à ne pas prédire les lignées non attendues).

*** La sensibilité évalue l'efficacité du modèle dans sa capacité à reconnaître les vrais positifs, (donc à prédire la lignée attendue).

Toutes les souches EAI6-BGD1 (n=11) sont prédites EAI5, ce qui est une proposition raisonnable, puisque EAI5 est la sous-lignée la plus basale du groupe EAI (Lazzarini et al., 2011).

2.2.3. Evaluation dans la complétude des informations de typage

Deux autres tests ont permis d'évaluer la pertinence de VNTR-TBbayes à prédire les valeurs manquantes relatives aux loci MIRU-VNTR. Un locus à faible variabilité, le MIRU4 (HGDI=0,36), est testé. Un autre locus, très variable, le MIRU40 (HGDI=0,70) est également testé. Dans 85% des cas, pour le MIRU4, la valeur la plus probable prédite est celle qui était attendue. Cette proportion est de 58% pour le MIRU40, Cependant, en utilisant le Z-score (comme décrit dans la partie Matériels et Méthodes), dans 97,33% des cas, la valeur attendue se retrouve dans l'intervalle de confiance pour MIRU4, et dans 98,84% des cas elle l'est pour le MIRU40,

2.3. Détermination par VNTR-TBbayes du profil MIRU-VNTR caractéristique de chaque lignée/sous-lignée phylogénétique

2.3.1. Présentation des profils MIRU-VNTR ou nœuds phylogénétiques centraux

Le spoligotype et le MIRU12, les plus probables ont été prédits pour 57 lignées/sous-lignées. Le spoligotypique le plus probable correspond presque toujours au prototype de la famille/sous-famille (Demay et al., 2012). Pour les sous-lignées : CAS1-Kili, EAI2-nonthaburi, H2, H37Rv, AFRI2/3, *M. canettii*, Manu3, PINI2, T2-uganda, T3-ETH, T4-CEU1, T5-RUS1 le spoligotype le plus probable ne correspondait pas au prototype.

2.1.1. Construction d'un arbre bayésien avec les différents nœuds centraux pour discuter des résultats de VNTR-TBbayes

Un arbre bayésien fut construit avec les 57 patterns MIRU12 (Figure 15), prédits par VNTR-TBbayes pour les 57 lignées/sous-lignées, pour discuter des résultats du test qui visait à prédire les lignées à partir du MIRU12 seul. Cette méthode a été choisie pour mettre en évidence des lignées spoligotypiques mal-définies et pour s'affranchir des phénomènes de convergence qui touchent le locus DR (Comas et al., 2009). L'arbre représente les liens entre les 57 lignées/sous-lignées, et la longueur des branches traduit la distance évolutive qui les sépare. Dans cet arbre, les lignées appartenant à *M. tuberculosis* au sens strict (Beijing, CAS, EAI, LAM, Haarlem, T, X, S) sont représentées en bleu afin de les distinguer des sous-espèces appartenant à MTC (*M. africanum*, *M. microti*, *M. pinnipedii*, *M. canettii*, *M. bovis*) représentées en vert.

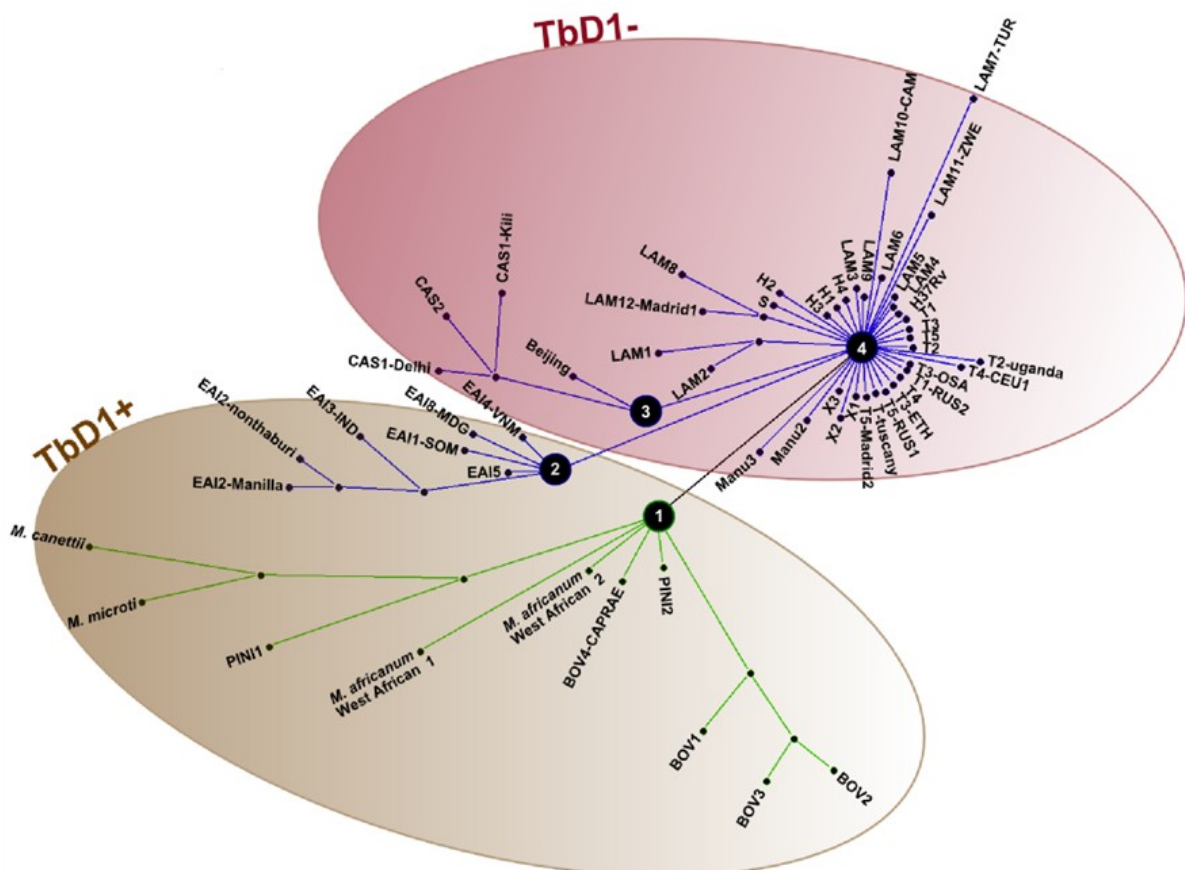


Figure 15. Liens phylogénétiques entre les lignées génotypiques du complexe *M. tuberculosis*. L'arbre bayésien fut construit avec le logiciel MrBayes3 (<http://mrbayes.csit.fsu.edu/>) avec les MIRU12 les plus probables pour chaque lignée selon VNTR-TBbayes.

L'arbre fait une séparation entre les lignées TbD1+ et ceux qui ne présentent pas la région TbD1 (TbD1-) (Brosch et al., 2002). L'arbre est divisé en 4 sous-arbres pour faciliter son interprétation. Les sous arbres 1 et 2 portent les sous-lignées TbD1+/PGG1, alors que les sous-arbres 3 et 4 portent respectivement celles du groupe TbD1-/PGG1 et du groupe TbD1-/PGG2-3 (à l'exception des lignées Manu).

Sous-arbre 1. Faisant partie du groupe TbD1+, *M. canettii* / *M. prototuberculosis* serait l'espèce la plus ancestrale (Gutierrez et al., 2005), suivie de *M. microti* and *M. pinnipedii* (PINI1). L'arbre bayésien montre que les représentants de ces deux dernières lignées partagent un ancêtre proche. Mais la lignée *M. microti* serait apparue avant *M. pinnipedii* PINI1. Cette observation ne corrobore pas les résultats précédents selon lesquels les signatures de ces lignées seraient définies respectivement par la perte de deux blocs d'espaceurs ; 1 à 36 et 39 à 43 pour *M. microti* et des trois blocs 1 à 3, 8 à 22, et 39 à 43 chez *M. pinnipedii* PINI1 (Brudey et al., 2006) (Demay et al., 2012). Vu que le locus DR évolue par la perte des espaceurs, ces signatures sous-entendent que la lignée *M. pinnipedii* serait apparue avant *M. microti*. Il est intéressant de constater que Smith et al. (2009), à l'issue d'une identification spécifique des souches *M. microti* par la perte de la région spécifique RD1^{mic}, la signature spoligotypique spécifique à cette lignée s'est caractérisée par les pertes 1 à 3, 8 à 22, et 39 à 43. Cette nouvelle signature de *M. microti* est confirmée par une étude de Panteix al. (Panteix et al., 2010). Ces deux lignées (*M. pinnipedii* PINI1 et *M. microti*) partagent donc la même signature spoligotypique. Cependant ces deux lignées ne sont pas pathogènes pour les mêmes hôtes, ce qui laisse penser qu'il s'agit bien de deux lignées distinctes. La lignée *M. pinnipedii* est retrouvée principalement chez les phoques (Cousins et al., 2003), alors que la lignée *M. microti* prédomine chez les petits rongeurs (Wells et al., 1937). Cette observation montre l'importance des loci MIRU-VNTRs dans la discrimination de ces deux lignées.

En suivant la chronologie de l'histoire évolutive inférée par le sous-arbre 1 (figure 15), *M. africanum* et *M. bovis* seraient les prochains à émerger. L'arbre bayésien suggère que *M. africanum* West African 1 (WA1) et *M. africanum* West African 2 (WA2) seraient proches de *M. bovis*, d'ailleurs ils partagent tous la perte de RD9 (Brosch et al., 2002). La chronologie évolutive suggère que WA1 serait apparu avant WA2, hypothèse corroborée par le fait que WA1 (Δ RD9) a moins de délétions que WA2 (Δ RD9, Δ RD7) (de Jong et al., 2010). *M. bovis* qui compte 4 sous-lignées

(BOV1 à BOV4) (Brudey et al., 2006). BOV1, 2 et 3 partagent un ancêtre commun, alors que BOV4-Caprae est plus proche de *M. africanum* WA2.

Sous-arbres 2 et 3. Ces deux arbres se partagent des souches du groupe PGG1. Cependant le sous-arbre 2 est exclusivement composé des sous-lignées ancestrales TbD1+/PGG1 alors que le sous-arbre 3 contient toutes les sous-lignées du groupe TbD1- /PGG1 (figure 15). Parmi toutes ces sous-lignées, EAI serait la plus ancestrale (Brosch et al., 2002), suivie par les lignées PGG1 CAS et Beijing plus récentes. La lignée CAS serait la première à émerger suivie de Beijing, avec qui elle partage un ancêtre proche.

Sous-arbre 4. Ce sous-arbre regroupe toutes les lignées TbD1- /PGG2/3 de MTC qui sont les lignées les plus récentes (LAM, Haarlem, X, S, T), et qui sont regroupées dans la grande famille Euro American de Gagneux et al. (Gagneux et al., 2007). Remarquons que toutes les sous-lignées évoluent à partir d'un ancêtre commun symbolisé par un noeud central noté 4 (figure 15). Il a été vu plus haut que VNTR-TBbayes est moins efficace dans la prédiction des lignées PGG2/3 que celles appartenant au groupe PGG1, (à partir des MIRU12 seuls) (tableau 3). On peut penser que le problème provient des MIRUs, mais il faut garder à l'esprit que la principale limite de ce système est qu'il soit basé sur les familles génotypiques étiquetés par des règles spoligotypiques, et qu'une mauvaise définition de ces règles au sein de la lignée Euro American ne peut être exclue.

Une étude récente, basée sur le polymorphisme des gènes impliqués dans les fonctions de réplication, de réparation et de recombinaison (3R) chez MTC, a souligné un problème de dénomination des lignées spoligotypiques modernes (Abadia et al., 2010). Par exemple les souches H1 et H2 partagent le même SNP *mgtC*^{545 (cgc-cac)} spécifique des souches Haarlem. Cependant ce SNP est absent chez toutes les souches H4, et présents chez certains isolats H3. De même VNTR-TBbayes, à partir des patterns MIRU12, a su prédire la sous-lignée attendue pour 100% des isolats H2, 66% des H3 et 65% des souches H1. Parmi les souches H3 mal classifiées avec VNTR-TBbayes, 52 % furent prédites X, 32% T and 16% LAM ou S. Ces résultats sont en accord avec le polymorphisme des 3R, qui révèle un lien (ou plutôt une confusion) entre H3, T2 et X (Abadia et al., 2010). Enfin, VNTR-TBbayes, avec les MIRUs, a considéré les souches T5-RUS1 comme des LAM, ces résultats corroborent les observations de Mokrousov et al., (2009).

LAM7-TUR présente la plus longue branche dans l'arbre bayésien, ce qui laisse supposer qu'il serait le premier à émerger parmi les PGG2/3. Les isolats LAM7-TUR se caractérisent par 2 copies d'IS6110, faible nombre de copies caractéristique des souches ancestrales. Cependant, LAM7-TUR partage un pattern MIRU12 avec T3-OSA (Zozio et al., 2005). Dans ce contexte, une étude récente a suggéré que LAM7-TUR n'appartiendrait pas au groupe phylogénétique LAM, puisque toutes les souches testées n'avaient pas le SNP *ligB*¹²¹² spécifique des LAM (Abadia et al., 2010). Alors que ces deux sous-lignées arborent des spoligotypes différents (SIT41 pour LAM7-TUR et SIT627 pour T3-OSA), les deux présentent le même pattern MIRU12 (MIT310) ainsi que le même SNPs sur *ligC*⁸⁰⁹ (tgg [trp] à ttg [leu]) (Abadia et al., 2010). Rappelons que VNTR-TBbayes a été construit en considérant LAM7-TUR comme une LAM et T3-OSA comme un T ce qui peut créer un biais dans la prédiction de ces sous-familles. Malgré cela, VNTR-TBbayes ne fait pas de discrimination entre les deux sous-familles et les place dans le groupe T. Parmi tous les isolats LAM, VNTR-TBbayes place en LAM toutes les souches appartenant à LAM1, LAM6 et LAM8. Les isolats appartenant à LAM2, LAM3, LAM4, LAM5, LAM9, LAM10-CAM, et LAM11-ZWE ne sont pas toujours prédits LAM. Par exemple, parmi les souches LAM3 (n=35), 48,6% sont prédites LAM (n=17), 31,4% T(n=11), 11,4% S (n=4) et les autres sont prédites H (n=2) et X (n=1). Dans ce contexte, la division du grand groupe phylogénétique LAM en deux sous-groupes par la délétion RD(Rio) (Gibson et al., 2008) a besoin d'être clarifiée grâce au SNP spécifique à LAM, aux LSPs, et aux minisatellites.

2.2. Evolution et perspectives de VNTR-TBbayes

2.2.1. Apport de nouvelles variables au modèle bayésien

2.2.1.1. Les variables potentielles

Les réseaux bayésiens démontrent, ici, leur capacité à extraire les informations cachée dans une base de données où l'exploitation de la grande quantité de données accumulées dépasse les capacités humaines. VNTR-TBbayes pousse à réévaluer le potentiel des minisatellites MIRUs. Ceux-ci soulèvent les mêmes problèmes déjà soulevés par les LSPs et les SNPs dans l'assignation des lignées. Une classification basée sur les MIRU-VNTR est impératives pour la confronter à celle basée sur les spoligotypes. Le schéma de classification spoligotypique est la plus discriminante, il est temps d'avoir une classification aussi

détaillée basée sur une autre méthode pour pouvoir les comparer et évaluer la pertinence de chacune à identifier les diverses lignées.

Une amélioration de VNTR-TBbayes passe nécessairement par une augmentation du nombre de variables considérées mais aussi par une bonne définition de ces variables. Ainsi il faut nécessairement une classification robuste au sein de MTC. Les variables qui peuvent être ajoutées portent sur la virulence ou les critères épidémiologiques. VNTR-TBbayes peut également être étendu à des variables relatives aux SNPs, aux LSPs, à la sensibilité aux antibiotiques et à des variables démographique et clinique relatives aux patients.

2.2.1.2. Etude réalisée au Nigéria : le statut VIH, une variable dans un modèle bayésien pour MTC ? (Article 7)

Le virus de l'immunodéficience humaine est un terrain propice au développement de la tuberculose-maladie. Plus de la moitié des individus coinfectés par le VIH et la tuberculose (TB) se trouve en Afrique subsaharienne qui regroupe à elle seule les deux tiers des personnes atteintes par le VIH/SIDA (syndrome de l'immunodéficience acquise). Avec une population de plus de 140 millions d'habitants, en 2007, le Nigeria fut classé au quatrième rang mondial parmi les pays les plus touchés par la tuberculose, avec une prévalence de 521 pour 100000 habitants (WHO report, 2009). Cette situation est aggravée par le VIH. En 2007, celui-ci avait une prévalence de 3,6% et le taux de coinfection TB-VIH était de 27% (UNAIDS, 2009). Bien que le résultat du traitement n'ait pas été évalué sur une forte proportion de patients en 2007, le taux de réussite du traitement était de 76%. La coinfection TB-VIH est également en cours d'extension, d'ailleurs 32% des cas de tuberculose sont dépistés à partir du VIH (WHO report, 2009). Entre 2005 et 2007, le taux de détection des nouveaux cas de tuberculose à frottis positifs, passa de 18% à 23%. Ce taux est donc resté trop faible. Autre point négatif : les informations disponibles relatives au génotypage de MTC sont limitées au Nigéria. De plus, dans les études génotypiques de MTC au Nigéria, la question de la coinfection TB-VIH n'est pas abordée. La présente étude, par conséquent, visait à étudier la structure de la population par spoligotypage chez des patients HIV+ et HIV- au Nigeria. Les résultats de typage furent ensuite comparés à la base de données SITVIT2.

2.2.1.2.1. Description de la collection d'échantillons de MTC de l'étude

La collection de souches fut répertoriée entre Avril 2004 et Octobre 2005 à Ibadan, dans l'état d'Oyo situé au Sud-Ouest du Nigéria (figure 18). Un total de 163 isolats furent collectés, ce qui correspond à 6% des cas de frottis positifs à Ibadan. Dans cet ensemble, on distingue 49 cas HIV+ et 114 HIV-. Les patients ont été recrutés à partir de différents centres de la stratégie DOTS à Ibadan. Toutes les souches ont été identifiées comme faisant partie de MTC, par la mise en culture et l'application des procédures microbiologiques standard au Laboratoire de recherche tuberculose du ministère de la Santé publique vétérinaire et de la médecine préventive, à l'Université d'Ibadan, à Ibadan au Nigeria (Cadmus et al., 2006).

2.2.1.2.2. Description de la population de MTC au Nigéria chez les patient HIV+

Chez les patients HIV+, le spoligotypage a produit 21 patterns différents pour les 49 isolats étudiés (tableaux 6 et 7). Sept d'entre eux sont des orphelins même après leur intégration dans la base de données SITVIT (tableau 7). Ainsi les 14 patterns restants constituent des SITs, autrement dit, ils sont présents plus de deux fois dans SITVIT. Ces SITs rassemblent les résultats de 42 patients. Parmi ces 14 SITs, seuls 4 étaient en cluster dans l'étude (n=32, 4 à 16 isolats par cluster). Les 17 isolats restant avaient des patterns orphelins parmi les HIV+. Les isolats en clusters dans ce groupe, appartenaient à deux clades : T1 (n=20/32, ou 62,5%) et LAM10-CAM (12/32, ou 37,5%). Cette dernière sous-famille phylogénétique présente une spécificité géographique pour le Cameroun et les pays voisins (Brudey et al., 2006), (Niobe-Eyangoh et al., 2004, 2003). Les lignées majeures circulant chez les personnes HIV+ suivent cet ordre d'occurrence : T1 (22/49, ou 44,9%); LAM10-CAM (13/49, ou 26,5%); AFRI (4/49, ou 8,2%); Manu, (3/49, ou 6,1%); et H3 (2/49, ou 4,1%). À l'exception des souches T (famille ubiquitaire), les patients HIV+ arborent davantage les spoligotypes SIT61 et SIT403 (LAM10-CAM) retrouvés en Afrique centrale et en Afrique de l'Ouest. Ils sont davantage partagés par les patients nigériens et béninois, voisins immédiats du Cameroun. Cependant, les souches Beijing qui représentent plus de 10% des souches du Bénin, ne sont pas retrouvées au Nigéria. Aucune souche Beijing n'est observée parmi les patients HIV+ du nigéria. Il en est d'ailleurs de même chez les HIV-.

2.2.1.2.3. Description de la population de MTC au Nigéria chez les patient HIV-

Chez les patients HIV-, le spoligotypage a produit 42 patterns différents pour les 114 isolats étudiés (tableaux 6 et 7). Parmi ces patterns, 8 d'entre eux sont des orphelins même après leur intégration dans la base de données SITVIT. Ainsi les 34 patterns restants sont des SITs. Ces SITs regroupent les résultats de 106 patients (tableau 6). Parmi les 34 SITs, 17 sont en cluster dans l'étude (n=89, 2 à 43 isolats par cluster). Les 25 isolats restant ont des patterns orphelins parmi les HIV- . Contrairement au groupe HIV+, les isolats en clusters dans ce groupe, appartiennent à plusieurs clades dont LAM9, LAM10-CAM, T1, T3, AFRI_1, AFRI_2, AFRI_3 et EAI5. LAM10-CAM regroupait la majorité des isolats (52/89, 58,43%). Les lignées retrouvées chez les patients HIV- sont : LAM10-CAM (59/114, ou 51,8%); AFRI (27/114, ou 23,7%); T1 (10/114, ou 8,8%); H3 (4/114, ou 3,5%); EAI (3/114, ou 2,6%); EAI5 (2/114, ou 1,8%); *M. bovis* (2/114, ou 1,8%); T3 (2/114, ou 1,8%); LAM9 (2/114, ou 1,8%); LAM6 (1/114, ou 0,9%); et LAM3 (1/114, ou 0,9%).

2.2.1.2.4. Conclusion sur le statut VIH et les profils génotypiques

LAM10-CAM est bien représenté chez les deux groupes avec une prévalence dans les deux cas du SIT61 (prototype de LAM10-CAM, 37,72% des cas HIV- et 12,24% des cas HIV+). Un des variant LAM10-CAM, SIT403 est observé dans 0,88% des cas dans le groupe HIV- alors qu'il l'est dans 12,25% des cas chez les HIV+. La famille T est également bien représentée, surtout par un variant SIT373 (n=16) chez les HIV+, qui cependant, est totalement absent de l'échantillon HIV-. Ceci laisse penser qu'il s'agirait d'un clone spécifique qui affecte près du 1/3 des patients coinfectés par la tuberculose et le VIH.

3. Une classification basée sur les données de typage MIRU-VNTR pour une phylogénie plus robuste de MTC (Article 8)

3.1. Définition d'un prototype MIRU-VNTR à l'aide d'une étude réalisée en Inde (Article 9)

3.1.1. Épidémiologie moléculaire des isolats MTC à Kerala en Inde à l'aide des techniques *IS6110-RFLP*, spoligotypage⁴³ et du typage par MIRU-VNTR

En Asie, l'Inde, la Chine, le Bangladesh, l'Indonésie et le Pakistan regroupent près de la moitié (48%) des nouveaux cas de tuberculose dans le monde (WHO Report, 2011). L'incidence globale de la tuberculose en Inde est de 2 millions de cas, ce qui représente 1/5 de l'incidence mondiale (avec 9,4 millions de cas), (Report Directorate of General of Health Services, 2011). Le poids de la tuberculose est certainement sous-estimé. En effet, de nombreux cas dans le pays ne seraient pas signalés en raison du manque d'accès aux établissements de santé, de la surpopulation, de la pauvreté et d'autres stigmatisations sociales. L'incidence élevée de la tuberculose en Inde est encore aggravée par l'émergence croissante de souches résistantes aux médicaments, y compris multi-résistante (MDR). Rappelons que les souches MDR ont une résistance à au moins deux des principaux antituberculeux, la rifampicine et l'isoniazide (Joseph et al., 2009), (Joseph et al., 2007), (Almeida et al., 2005).

Les souches MTC ont 99,9% de similarité entre elles (Boddinghaus et al., 1990), Sreevatsan et al., 1997). Il y a des preuves croissantes sur le fait que des différences génétiques spécifiques soient associées à des locations géographiques (Brudey et al., 2006, Demay et al., 2012). Seules quelques études ont été menées en Inde pour caractériser les géotypes de MTC (Das et al., 1995), (Radhakrishnan et al., 2001), (Siddiqi et al., 2001), (Mistry et al., 2002), (Varma-Basil et al., 2011), Gutierrez et al. (2006). Parmi ces études, une seule fut réalisée à Kerala. Celle-ci a fait appel seulement à l'*IS6110-RFLP*, méthode non appropriée pour les régions présentant des souches MTC à faible nombre de copies d'*IS6110* (Radhakrishnan et al., 2001). En effet, Kerala est un état du Sud-Ouest indien. Cette zone géographique est l'une des zones de prédilection de la lignée EAI qui se caractérise par un faible nombre de copies d'*IS6110*,



Figure 16. États du Sud de l'Inde. Kerala est représenté en vert.

3.1.1.1. Description de la collection de souches de Kerala en Inde

Cette étude inclut 168 isolats de MTC de patients suivis dans différentes cliniques de Kerala en Inde. Les isolats proviennent de 129 hommes et 39 femmes ayant entre 18 et 78 ans. Ces isolats étaient présents dans la collection du laboratoire qui a été analysée pour évaluer la résistance des souches aux antibiotiques. Les échantillons de crachats ont été recueillis auprès des patients le matin, transportés au laboratoire sur de la glace puis traités le même jour. L'expectoration a été traitée par la méthode modifiée de Petroff et les bactéries ont été cultivées sur un milieu solide de Löwenstein-Jensen (Petroff, 1915). La morphologie et la propriété acido-alcool résistante ont été vérifiées chez les colonies. Les isolats ont été caractérisés par des techniques biochimiques standard. Les isolats ayant répondu positivement aux tests biochimiques ont été confirmés comme *M. tuberculosis* après une PCR (Anilkumar et al., 2012). Trois différentes méthodes de génotypage ont été utilisées : IS6110-RFLP sur 156 échantillons, le spoligotypage et le typage par MIRU-VNTR (sur 15 loci dont les 12 loci standards et 3 ETRs ; ETR-A, ETR-B, ETR-C) sur l'ensemble des 168 échantillons.

3.1.1.2. Description du paysage génotypique à Kerala en Inde

3.1.1.2.1. Description des spoligotypes répertoriés

Le spoligotypage des 168 isolats indiens a produit 68 spoligotypes différents. Parmi eux, on compte 17 orphelins (10,11%) (tableaux 8 et 9). La comparaison avec

SITVIT (qui contient 86 000 souches), révèle que ces 17 isolats de Kerala ne sont retrouvés nulle part ailleurs. Les 51 patterns restants correspondent donc à des SITs dont 42 (regroupant 136 isolats, 80,95%) étaient déjà répertoriés dans la base. Ainsi, 9 SITs (rassemblant 15 isolats, 8,92%) ont été nouvellement créés. Enfin, 21 patterns formaient déjà des clusters (de 2 à 48 isolats) à l'intérieur de l'échantillon de l'étude par opposition aux 30 patterns uniques de l'étude. Le plus grand cluster est celui du SIT11 avec 48 isolats (28,6% des isolats). SIT11 est le prototype de EAI3-IND qui présente une forte géospécificité pour l'Inde (Brudey et al., 2006). La majorité des isolats (108/168, 64,28%) appartiennent à la grande lignée EAI. La Distribution des isolats dans les sous-lignées EAI se fait selon cet ordre : EAI3-IND (60/168), EAI1-SOM (21/168), EAI5 (22/168), EAI6-BGD1 (3/168), EAI-MDG (1/168) et EAI7-BGD2 (1/168). Les autres lignées/sous-lignées les mieux représentées sont : CAS1-Delhi (4.16%) et Beijing (2.4%). La distribution géographique mondiale des patterns formant des clusters ayant plus de 2 isolats, met en évidence la spécificité phylogéographique de presque toutes les sous-lignées EAI de cette étude pour l'Asie du Sud. Parmi les 17 orphelins, 11 patterns appartiennent à la lignée EAI. D'autres clades sont répertoriés : T3 (1/168), Manu 1 (2/168), LAM9 (1/168), T1 (5/168), H3 (1/168), H1 (1/168). Le pouvoir discriminant du spoligotypage atteint 0,9113 dans l'échantillon de Kerala.

étude menée à Kanpur (ville de l'état de l'Uttar Pradesh au Nord de l'Inde), le MIRU26 était le plus variable (Sharma et al., 2008).

VNTR	Nombre de répétitions											HGDI	Rang
	0	1	2	3	4	5	6	7	8	9	10		
MIRU 2	0	2	165	1	0	0	0	0	0	0	0	0.0354	12
MIRU 4	1	1	37	1	8	100	5	4	2	8	1	0.5945	4
MIRU 10	0	0	0	18	125	16	9	0	0	0	0	0.4255	7
MIRU 16	0	4	4	147	10	1	2	0	0	0	0	0.2309	9
MIRU 20	0	5	162	1	0	0	0	0	0	0	0	0.0696	11
MIRU 23	0	0	1	3	5	29	118	5	3	0	4	0.4767	5
MIRU 24	2	40	120	6	0	0	0	0	0	0	0	0.4343	6
MIRU 26	0	0	131	0	5	18	3	9	0	2	0	0.3785	8
* MIRU 27	0	1	3	159	4	0	0	0	0	0	0	0.0931	10
^ MIRU 31	0	0	2	24	67	67	2	3	0	0	0	0.6524	2
# MIRU 39	0	37	26	89	10	2	0	0	0	0	0	0.6294	3
MIRU 40	0	5	19	66	71	7	0	0	0	0	0	0.6555	1
ETR A	0	1	8	19	24	6	78	3	16	11	2	0.7382	1
ETR B	0	77	46	0	35	3	5	1	1	0	0	0.6743	2
ETR C	0	8	31	29	95	5	0	0	0	0	0	0.6169	3

Tableau 10. Diversité allélique des minisatellites VNTR parmi les isolats MTC de Kerala en Inde. Le signe « \$ » désigne un nombre total d'isolats de 167 vu que un présentait des bandes multiples. Le signe « * » désigne un nombre total d'isolats de 165 vu que 3 présentaient des bandes multiples. Le signe « # » désigne un nombre total d'isolats de 164 vu que 4 présentaient des bandes multiples.

3.1.2. Illustration de l'importance des profils MIRU-VNTR centraux prédits par VNTR-TBbayes

Afin de souligner l'importance de ces nœuds centraux prédits par VNTR-TBbayes pour chaque sous-lignée, un arbre MST a été construit avec les 164 MIRU12 de l'échantillon de Kerala (Inde) (figure 17). VNTR-TBbayes a montré une bonne capacité à prédire les lignées à partir des MIRU12 seuls chez les souches PGG1, ce qui présage une bonne définition des lignées spoligotypiques appartenant au groupe PGG1. L'échantillon d'Inde où les souches EAI (appartenant au groupe PGG1) sont bien représentées, est donc une collection de souches idéale pour comprendre le lien entre MIRU12 et familles spoligotypiques.

Sur l'arbre de la figure 17, des cercles de couleur entourent les profils appartenant à la même sous-lignée spoligotypique (Brudey et al., 2006). Par exemple, les clusters de souches que forment les sous-lignées EAI sont encadrés en utilisant diverses déclinaisons de rouge.

Cet arbre (figure 17) peut être scindé en 3 parties relatives à la subdivision du complexe *M. tuberculosis* en 3 groupes grâce aux SNPs *katG* et *gyrA* (Sreevatsan et al. 1997) et au LSP *TbD1* (Brosch et al. 2002). Le premier groupe PGG1/*TbD1*+ se compose des diverses sous-lignées ancestrales EAI. Le second groupe PGG1/*TbD1*- se compose des souches CAS et Beijing alors que le troisième groupe se compose des sous-lignées modernes (PGG2-3/*TbD1*-). D'après l'arbre (figure 18) *Manu* ferait partie du troisième groupe. L'arbre MST montre les noeuds centraux de EAI1-SOM, EAI3-IND, Beijing et CAS1-Delhi qui sont respectivement MIT64, MIT69, MIT17 and MIT318. Ces patterns de MIRU12 correspondent aux noeuds centraux prédits par VNTR-TBbayes pour chacun de ces sous-lignées. Cet arbre montre que tous les membres d'une sous-lignée sont réunis autour d'un pattern central. Cette agglomération inclut tous les représentants de la sous-lignée, et le pattern central est le plus commun et est également celui qui génère sans doute le plus de variants.

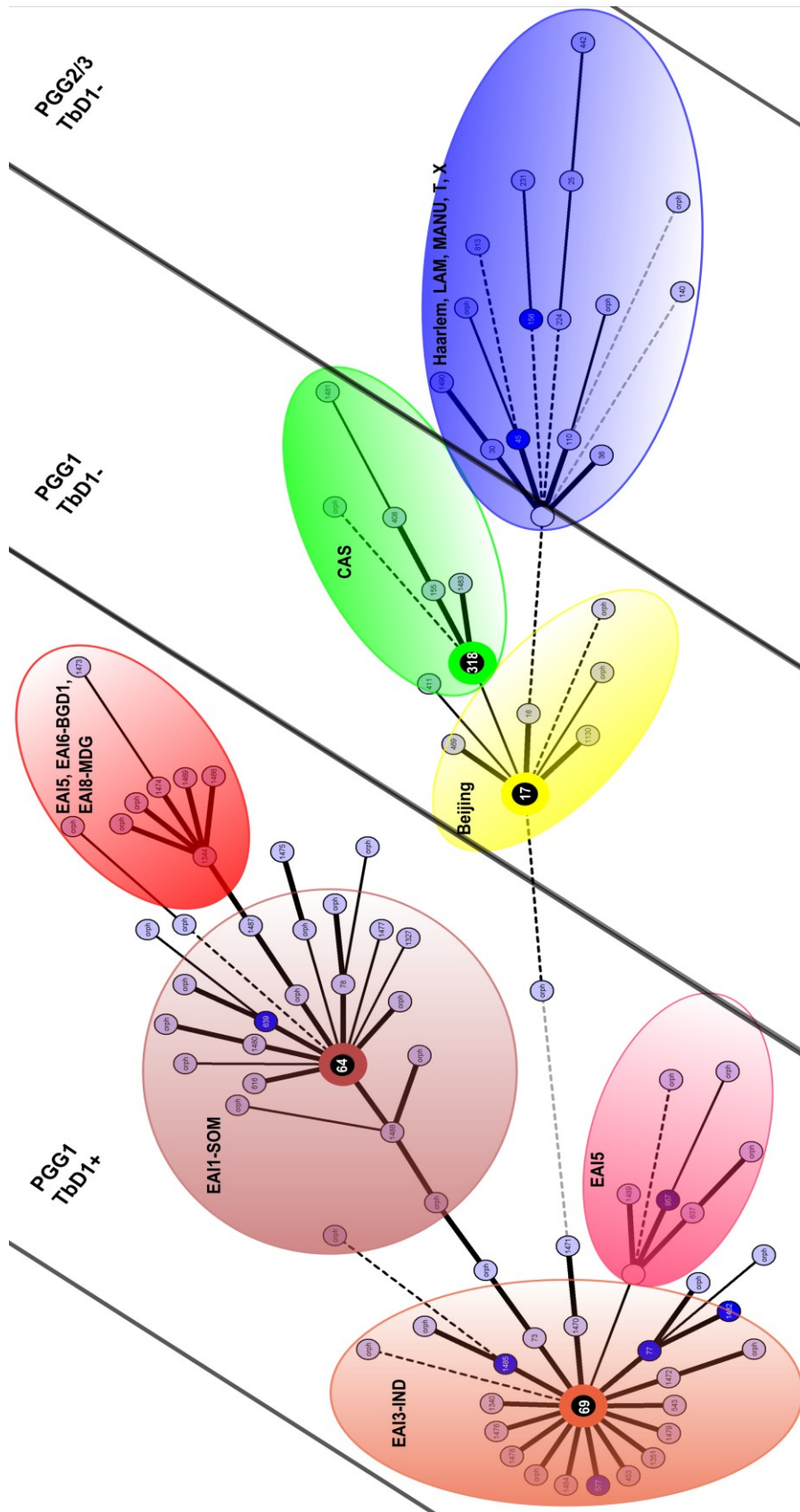


Figure 17. Arbre phylogénétique fait avec 164 souches de Kerala en Inde. Cet arbre est basé sur les 12 loci MIRU-VNTR, et est construit selon la méthode du Minimum Spanning Tree (MST) avec le logiciel BioNumerics. L'insertion de nœuds hypothétiques a été autorisée et ceux-ci sont représentés par des cercles vides (nœuds sans numéro).

3.1.3. Méthodologie utilisée pour l'identification des lignées selon les minisatellites MIRU-VNTR et la classification des patterns MIRU12

Identifier les lignées selon les minisatellites MIRU-VNTR, revient à identifier tous les nœuds centraux de MTC, et de les considérer comme des prototypes définissant chacun une lignée. Pour arriver à cette fin, 7009 isolats de la base SITVIT pour lesquels les résultats de MIRU12 et de spoligotypage sont disponibles, ont formé un échantillon pour la construction d'un arbre MST. Cet échantillon a d'abord servi à l'identification des diverses lignées et sous-lignées selon les 12 loci MIRU-VNTR, puis ensuite, à la représentation graphique de la distribution de celles-ci dans les différentes sous-régions du monde. L'échantillon comprenait seulement des patterns MIRU12 présents au moins deux fois dans la base SITVIT.

Après l'identification des lignées VNTR, un algorithme de classification fut conçu afin d'attribuer à un pattern MIRU12 une des familles nouvellement définies. Pour classer un pattern MIRU-VNTR, l'algorithme de classification le compare à tous les prototypes identifiés. Pour cela, il passe par 2 étapes. Premièrement, il retient le (s) prototype (s) avec lequel (lesquels) le pattern MIRU12 non classifié partage le plus de valeurs de loci. Si plus d'un prototype est retenu, l'algorithme passe à la deuxième étape. Il calcule, pour chaque prototype retenu, un Z-Score Cumulé (ou Cumulative Z-score, CZS), afin de déterminer le prototype le plus proche du pattern à classer. Le prototype avec lequel le pattern à classer a le CZS le plus bas, dévoile la lignée d'appartenance. Le CZS, formule utilisée pour évaluer la distance entre un pattern MIRU-VNTR quelconque et un prototype, est décrit dans la partie « Matériels et méthodes ».

3.2. Apport de nouvelles données de typage pour compléter la base SITVIT

Des données issues des études menées, dans le cadre de cette thèse, font partie des 7009 isolats utilisés pour l'identification des prototypes MIRU-VNTR. Il était nécessaire d'apporter de nouveaux profils, décrits simultanément par le spoligotypage et le MIRU12, issus de certaines régions pas bien représentées dans la base SITVIT. Ce renforcement de la représentativité de toutes les régions du monde, vis-à-vis de MTC, vise à bien déterminer les nœuds centraux et à évaluer les paramètres du CZS, pour chaque locus VNTR dans chaque lignée/sous-lignée. En effet, la moyenne et l'écart type du nombre de répétition, de chaque locus à l'intérieur

de chaque lignée/sous-lignée, sont les deux paramètres indispensables au calcul du CZS (voir la partie concernée dans « Matériels et Méthodes »).

3.2.1. Profils génotypiques caractéristiques de la côte Ouest africaine dans une étude menée au Nigéria (Article 10)

3.2.1.1. Description de la collection de souches

L'étude fut menée dans l'état de Cross River, localisé dans le Sud du Nigéria (figure 18). Un total de 137 frottis positifs fut collecté sur une période de 12 mois, entre juin 2008 et mai 2009. Les patients étaient âgés de 15 à 64 ans. En 2009, 44863 frottis positifs ont été réalisés, pour une population globale de 155 millions d'habitants au Nigéria. L'état de Cross River compte 2,89 millions d'habitants soit 1,86% de la population globale. Ainsi, pour l'état du Cross River, on s'attendait à obtenir 835 frottis positifs mais l'on a répertorié que 16,4% de ce nombre.



Figure 18. États d'Oyo et du Cross River au Nigéria

Parmi les 137 frottis, il y avait 97 isolats purs de mycobactéries dont 81 appartenant à MTC. Ceux-ci ont été spoligotypés et typés selon le format des 12 loci MIRU-VNTR.

3.2.1.2. Description des profils génotypiques au Nigéria

Les 81 isolats cliniques ont produit 27 spoligotypes différents. Dix patterns étaient déjà en clusters dans l'étude (avec 2 à 32 isolats par cluster). Dix-sept SITs regroupant 64 isolats existaient déjà dans SITVIT, alors que 4 SITs ont été nouvellement créés. Les 6 spoligotypes restants sont restés orphelins.

3.2.3. Présentation et analyse des nouvelles lignées basées sur 12 loci MIRU-VNTR

3.2.3.1. Présentation des 41 signatures génotypiques identifiées

L'arbre MST, réalisé avec les 7009 isolats de SITVIT et basé sur les minisatellites MIRU-VNTR, a mis en évidence 41 nœuds centraux qui correspondent à 41 sous-lignées. De cet arbre ressort 7 nœuds centraux majeurs (tableau 13) représentés par les MITs : 57, 17, 116, 934, 664, 49 et 60, Le MIT57 regroupe plusieurs sous-nœuds : 56, 59, 64, 69 et 57 lui-même. Le MIT17 comprend 16, 83, 86, 93, 99, 101, 68, 261 ainsi que le 17 lui-même. La plus grande lignée a pour nœud central le MIT116 et comprend 23 sous-nœuds : 7, 8, 12, 15, 25, 33, 34, 40, 42, 43, 45, 46, 112, 121, 125, 128, 163, 190, 212, 213, 224, 246 et le 116 lui-même. Il a été impossible de faire une image de l'arbre MST établissant des relations phylogénétiques entre les 7009 isolats. En effet, la grande quantité d'isolats rendait l'image illisible. Pour pallier à ce problème, un autre arbre fut refait avec les 41 MITs identifiés comme étant des nœuds centraux (figure 19).

Lignée basée sur MIRU12					Lignée de Gagneux (Gagneux and Small 2007)	Lignée de Brudey dans SpoIDB4/SITVIT (Brudey et al 2006)		
Lignée	Nœud central (MIT)	sous-lignée	Nœud central de la sous-lignée	MIRU12	Lignée	Lignée	Absence d'espaces caractéristiques	SIT
Indo-Oceanic	57	Indo-Oceanic-57	57	254326223533	Indo-Oceanic	East African-Indian (EAL)	29-32,34	236
		Indo-Oceanic-56	56	254326223432				
		Indo-Oceanic-59	59	264225223533				
		Indo-Oceanic-64	64	254326223513				
		Indo-Oceanic-69	69	254326223434				
East Asian and African Indian	17	East Asian-17	17	223325173533	East Asian	Beijing	1-34	1
		East Asian-16	16	223325153533				
		East Asian-83	83	223325163533				
		East Asian-86	86	223325173433				
		East Asian-93	93	223425173533				
		East Asian-99	99	223325173543				
		East Asian-101	101	223325173523				
		East-African Indian-68	68	225425173533	East-African Indian	Central-Asian (CAS)	4-7,23-34	26
		East-African Indian-261	261	227425113434				
Euro American	116	Euro American-116	116	223325153323	Euro American	Haarlem, Latin American and Mediterranean (LAM), X, T, S	33-36	53
		Euro American-7	7	222325153323				
		Euro American-8	8	223125153324				
		Euro American-12	12	223315153323				
		Euro American-15	15	223325153322				
		Euro American-25	25	224226153321				
		Euro American-33	33	224325153323				
		Euro American-34	34	224325153324				
		Euro American-40	40	225125113322				
		Euro American-42	42	225313153323				
		Euro American-43	43	225323153323				
		Euro American-45	45	225325153323				
		Euro American-46	46	225325153324				
		Euro American-112	112	223325143324				
		Euro American-121	121	223325143323				
		Euro American-125	125	223325153324				
		Euro American-128	128	223226153321				
		Euro American-163	163	224126152321				
		Euro American-190	190	124326153220				
		Euro American-212	212	233325153324				
Euro American-213	213	224326153323						
Euro American-224	224	223326153321						
Euro American-246	246	124326153324						
West African I	934	West African I	934	224424244221	West African lineage I	AFRI2, AFRI3	8-12, 37-39	438
West African II	664	West African II	664	236424253522	West African lineage II	AFRI1	7-9, 39	181
						BOV_4-CAPRAE	1,3, 16, 28, 39-43	647
<i>M. bovis</i>	49	<i>M. bovis</i>	49	232324253322		<i>M. bovis</i>	3, 16, 39-43	3090
<i>M. canettii</i>	60	<i>M. canettii</i>	60	323212632428		CANETTII	30, 36	592

Tableau 13. Description des 7 lignées majeures et 41 sous-lignées (de MTC) basées sur les 12 loci MIRU-VNTR. Ces lignées/sous-lignées ont été identifiées à l'aide de l'arbre MST construit avec les 7009 souches de la base SITVIT.

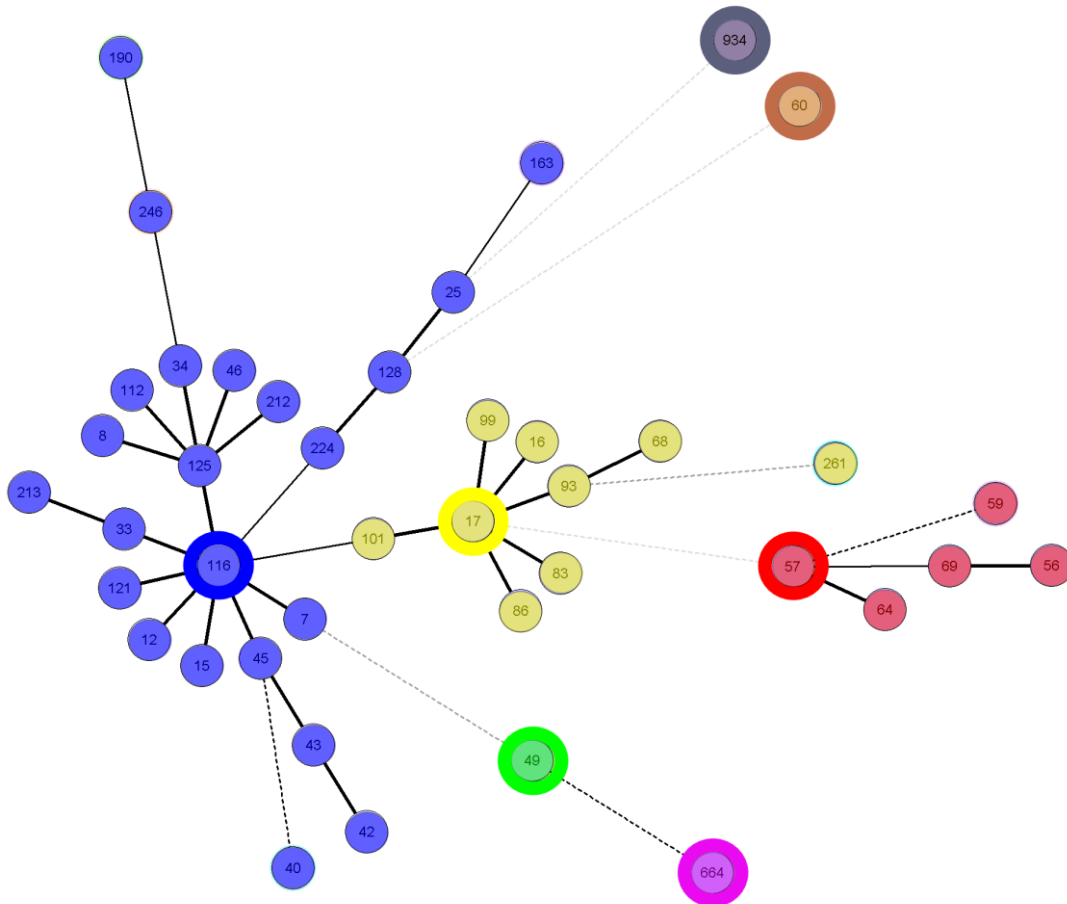


Figure 19. Arbre phylogénétique MST réalisé avec les 41 nœuds centraux qui sont les prototypes des différentes sous-lignées MIRU-VNTR. Arbre construit avec le logiciel BioNumerics.

3.2.3.2. Confrontation de cette nouvelle classification aux autres schémas de classification publiés

Une confrontation de cette nouvelle classification aux autres déjà établies, était nécessaire pour pouvoir attribuer un nom aux différentes sous-lignées identifiées. Cette comparaison a été faite avec la classification de Brudey (Brudey et al. 2006), en reclassifiant les 7009 souches de SITVIT. Les isolats de la base de données en ligne MIRU-VNTR_{plus} (<http://www.miru-vntrplus.org/MIRU/index.faces>) ont été également reclassifiés, dans le but de comparer la classification MIRU-VNTR et celle de Gagneux (Gagneux et al., 2007). Dans la base MIRU-VNTR_{plus}, les isolats *M. microti* ont été écartés, car cette espèce de MTC n'était pas présente parmi les 7009 souches utilisées pour identifier les prototypes MIRU-VNTR. Ainsi 176 isolats de MIRU-VNTR_{plus} ont été reclassifiés.

3.2.3.2.1. Reclassification des profils de la base de données MIRU-VNTRplus

La reclassification des 176 isolats de MIRU-VNTR $plus$ est présentée dans l'annexe 5. Comme le montre l'annexe, les isolats Indo-Oceanic (d'après la classification de Gagneux) sont dans les différentes sous-familles de la lignée ayant pour prototype le MIT57. Ceux appartenant à la lignée East African Indian et à la lignée East Asian, ont été classifiés dans les sous-lignées de la lignée définie par MIT17. Les souches de la lignée moderne Euro American se répartissent dans les différents sous-groupes de la lignée étiquetée par le prototype MIT116. Les souches *M. bovis* correspondent à la lignée MIT49. Les souches *M. caprae* et West African-2 composent la lignée définie par MIT664. On ne peut expliquer clairement pourquoi ces deux lignées ne forment qu'une lignée selon la nouvelle classification, mais on sait que AFRI1 partage avec toutes les espèces de MTC qui sont pathogènes aux animaux (dont BOV_4-CAPRAE) les régions de différence RD9, RD7, RD8, RD10, ainsi qu'une variation de 6 bp dans le gène *pks* (Brosch et al., 2002).

Une souche ayant pour ID 9550/00 dans la base MIRU-VNTR $plus$, est classifiée à la fois dans la lignée du MIT664 et celle du MIT49 (annexe 5). Un arbre MST réalisé avec les 176 isolats de MIRU-VNTR $plus$ montre que les lignées définies par les nœuds centraux MIT934, MIT664 et MIT49 sont proches phylogénétiquement (figure 20). Peut-être que le pattern MIRU12 de la souche 9550/00 est un intermédiaire entre les deux lignées du MIT664 et du MIT49.

Les souches West African 1, et *M. canettii* sont respectivement dans les lignées ayant pour prototype le MIT934 et le MIT60 (annexe 5).

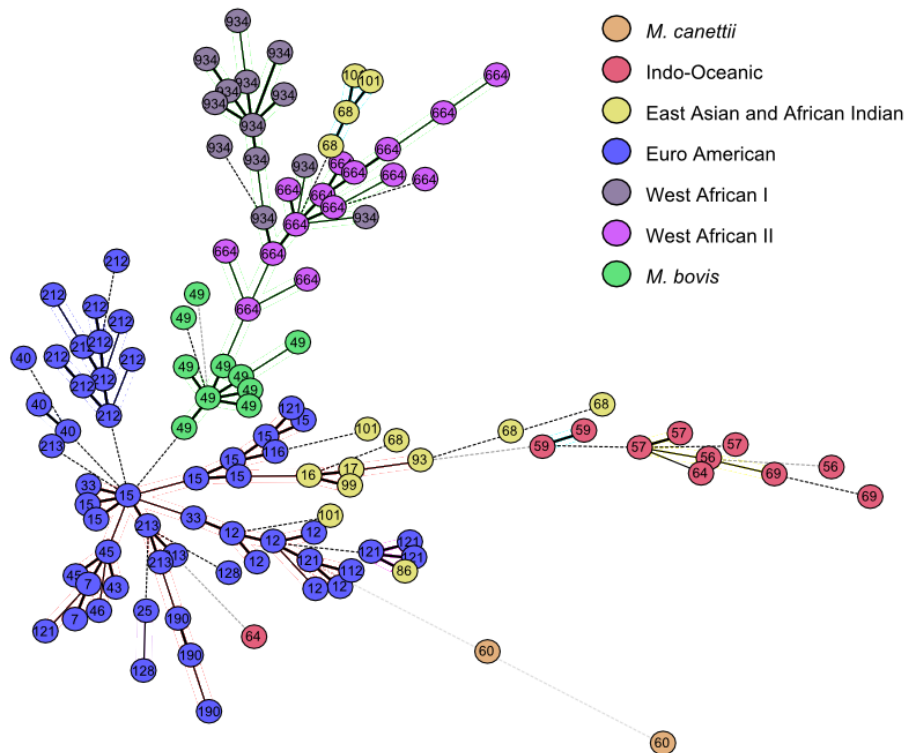


Figure 20. Arbre MST réalisé avec les 12 loci MIRU-VNTR de 176 isolats de la base de données en ligne MIRU-VNTRplus (<http://www.miru-vntrplus.org/MIRU/index.faces>)

3.2.3.2.2. Reclassification des profils de la base de données SITVIT

La reclassification des 7009 souches de SITVIT a permis de faire les mêmes correspondances avec la classification de Brudey qu’avec celle de Gagneux. Par exemple, EAI correspond à la lignée MIT57, Beijing et CAS correspondent à la lignée définie par MIT17. En poussant l’analyse plus loin, on s’aperçoit que les sous-lignées PGG1 de Brudey, plus précisément celles de EAI, CAS et Beijing, ont des correspondances avec les sous-lignées basées sur les loci MIRU-VNTR. Par exemple, 68,6% des souches EAI1-SOM correspondent à la sous-lignée définie par le MIT64, 72,97% des isolats EAI3-IND correspondent à la sous-lignée MIT69, 77,14% des souches EAI2-Manilla sont de la sous-lignée MIT56 (tableau 15). De même à l’intérieur du clade défini par le MIT17, 68,07% des isolats CAS1-Delhi sont dans la sous-lignée du prototype MIT68 et 74,19% des isolats CAS1-Kili sont dans celui du prototype MIT261 (tableau 14). Les isolats Beijing sont classifiés dans les autres sous-lignées de la lignée MIT17, c’est-à-dire celles définies par les patterns MIT17, MIT16, MIT83, MIT86, MIT93, MIT99, et MIT101. Le pourcentage de Beijing

retrouvé dans ces sous-lignées MIRU-VNTR varie entre 63% et 100%. Les souches AFRI2 et AFRI3 sont dans la lignée ayant pour prototype MIT934. Les isolats AFRI1 et BOV_4-CAPRAE sont dans la lignée du prototype MIT664. Les souches *M. bovis* sont dans la lignée du MIT49. Celles de *M. canettii* se révèlent être des souches de la lignée du MIT60,

Lignée MIRU-VNTR	Node central (MIT)	Sous-lignée MIRU-VNTR	Lignée majeure (Brudey et al., 2006) associée	Sous-lignée majeure (Brudey et al., 2006) associée
Indo-Oceanic	57	Indo-Oceanic-57	98,53 % EAI	42,65% EAI5
		Indo-Oceanic-56	97,14% EAI	77,14% EAI2-Manila
		Indo-Oceanic-59	100% EAI	57,5% EAI5
		Indo-Oceanic-64	97,67% EAI	68,60% EAI1-SOM
		Indo-Oceanic-69	97,37% EAI	72,97% EAI3-IND
East Asian and African Indian	17	East Asian-17	97,64% Beijing	
		East Asian-16	97,86% Beijing	
		East Asian-83	99,01% Beijing	
		East Asian-86	96,72% Beijing	
		East Asian-93	85,71% Beijing	
		East Asian-99	100% Beijing	
		East Asian-101	63,33% Beijing	
		East-African Indian-68	94,12% CAS	68,07% CAS1-Delhi
		East-African Indian-261	87,10% CAS	74,19% CAS1-Kili

Tableau 15. Comparaison de la nouvelle classification basée sur les 12 loci MIRU-VNTR avec celle de Brudey (Brudey et al., 2006). Dans ce tableau, seules deux lignées MIRU-VNTR sont considérées.

Les souches modernes PGG2/3 des lignées LAM, Haarlem, T, X, S sont réparties dans les différents sous-groupes de la lignée du prototype MIT116. Certaines correspondances entre la classification de Brudey et la nouvelle sont observées. LAM10-CAM correspond à la sous-lignée du MIT12. T2-Uganda correspond à celle du MIT121. LAM11-ZWE correspond à la sous-lignée MIT163. T5-RUS1 correspond à celle de MIT190, S correspond à la sous-lignée définie par le prototype MIT212. LAM12-Madrid1 est associée à celle du MIT213. T5-Madrid2 est associée à la sous-lignée ayant pour prototype MIRU-VNTR MIT224. LAM7-TUR, T3-ETH et T3-OSA correspondent à la sous-lignée du MIT40,

Il est difficile de faire d'autres correspondances au sein des lignées/sous-lignées modernes PGG2/3. Ce qui nous amène à la question : comment expliquer les discordances entre la classification de Brudey et celle nouvellement établie avec les

12 loci MIRU-VNTR. Des éléments de réponse vont être apportés à cette question dans le paragraphe 4.2.1 de ce chapitre.

3.2.3.3. Superposition de cette nouvelle classification à la classification de Gagneux basée sur les LSPs

Vu que cette nouvelle classification basée sur les minisatellites MIRU-VNTR se superpose bien à celle de Gagneux laquelle est basée sur les LSP, nous avons choisi de dénommer les lignées/sous-lignées en ajoutant le MIT prototype au nom de la lignée de Gagneux. Par exemple, la lignée définie par le MIT212 est appelée “Euro American-212” (et fait référence à la lignée S de Brudey). Un autre exemple : la lignée définie par le MIT69 est nommée “Indo Oceanic-69” (et fait référence à la lignée EAI3-IND de Brudey). Cependant il est intéressant de remarquer que deux lignées distinctes de Gagneux « East African Indian » et « East Asian » forment un seul grand groupe, nommé « East Asian and African Indian » (ou EAAI). La liste des noms des lignées/sous-lignées de cette nouvelle classification est dans le tableau 13.

3.2.4. Hypothèse évolutive et relations phylogénétiques entre les lignées nouvelles définies chez MTC

3.2.4.1. Construction d'une phylogénie bayésienne

Pour valider ou rejeter le fait que deux lignées distinctes de Gagneux « East African Indian » et « East Asian » forment un seul groupe phylogénétique, un arbre a été construit selon une méthode bayésienne, méthode différente de celle du MST. Cet arbre a été construit avec les 37 nœuds centraux identifiés au sein de *M. tuberculosis* au sens strict (figure 21). Les trois grands groupes, PGG1/TbD1+ (en rouge, Indo-Oceanic), PGG1/TbD1- (en jaune, East Asian and African Indian), PGG2-3/TbD1- (en bleu, Euro American) sont bien séparés sur cet arbre.

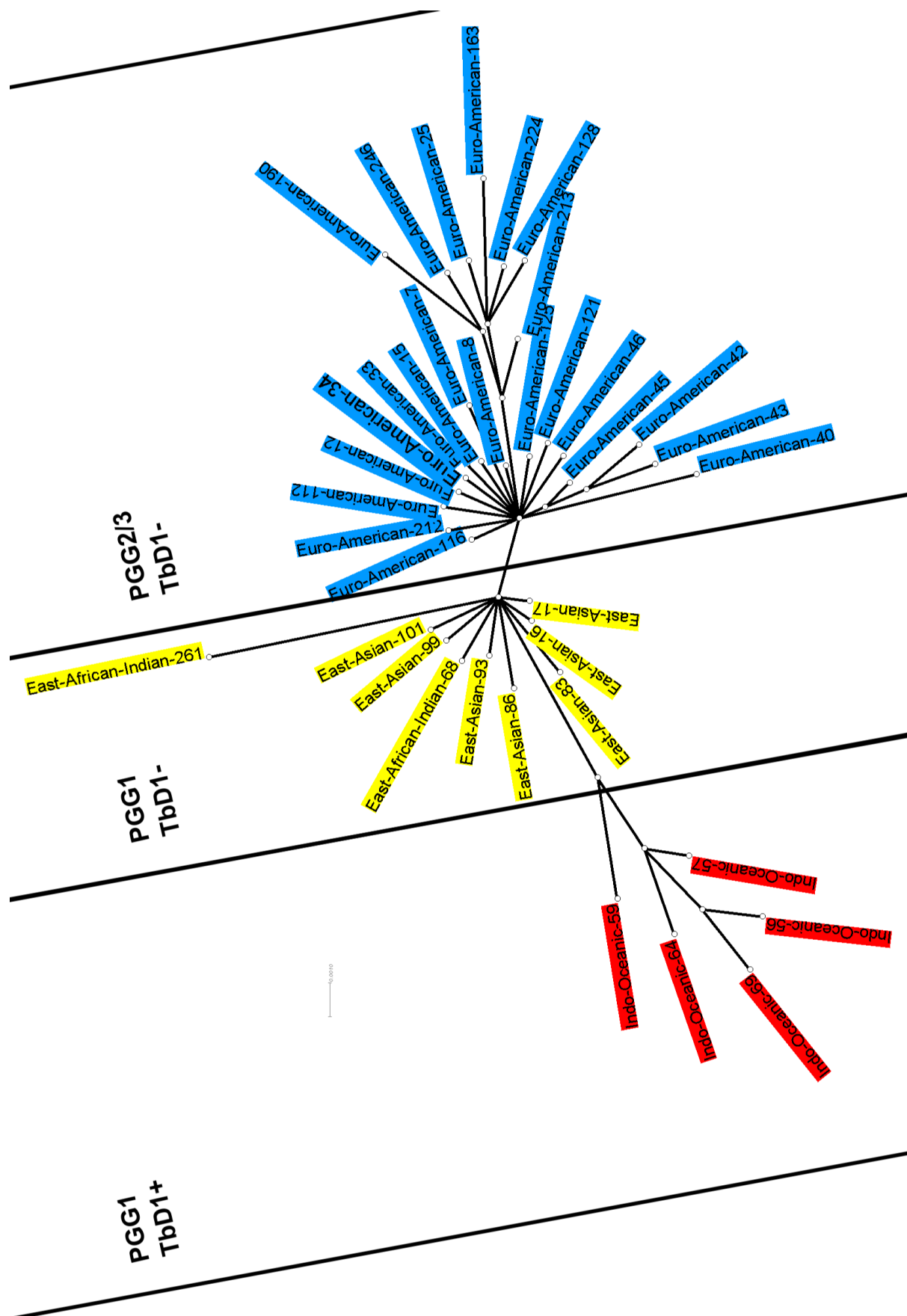


Figure 21. Liens phylogénétiques entre les sous-lignées MIRU-VNTR définies par les 37 nœuds centraux identifiés au sein de *M. tuberculosis* au sens strict. Cet arbre bayésien fut construit avec le logiciel MrBayes3 (<http://mrbayes.csit.fsu.edu/>).

Dans cet arbre bayésien, les sous-lignées East-African Indian-68 and East-African Indian-261 (correspondant également aux deux sous-lignées de CAS), partagent un noeud central avec les autres sous-lignées : East Asian-17, East Asian-16, East Asian-83, East Asian-86, East Asian-93, East Asian-99 and East Asian-101 (ces dernières correspondent à des subdivisions de la lignée Beijing). Ceci a donc permis de maintenir le nom “East Asian and African Indian” pour la lignée définie par le noeud central MIT17.

Le fait que la méthode bayésienne corrobore les trois grands groupes phylogénétiques (Indo-Oceanic, EAAI et Euro American), trouvés par la méthode MST au sein de *M. tuberculosis*, donne plus de crédit au nouveau schéma de classification MIRU-VNTR.

3.2.4.2. Identification de deux groupes phylogénétiques très importants chez les souches modernes PGG2/3

L'arbre bayésien montre deux branches phylogénétiques distinctes qui portent plusieurs sous-lignées de Euro American (MIT116). La première branche porte les sous-lignées Euro American-45, Euro American-43 et Euro American-42. La seconde porte les sous-lignées Euro American-213, Euro American-190, Euro American-246, Euro American-25, Euro American-163, Euro American-224 et Euro American-128. Ces deux branches ont été également observées sur l'arbre MST construit avec les 7009 isolats de SITVIT.

Après la reclassification de l'échantillon de 7009 isolats de SITVIT, 92% des souches Haarlem se retrouvèrent dans les 3 sous-lignées : Euro American-45, Euro American-43 et Euro American-42. Ces trois dernières sous-familles ont donc été compilées en un groupe nommé : *Haarlem-42/43/45*. La reclassification a aussi permis de constater que 74,33% des isolats LAM et 100% des isolats T5-RUS1 se retrouvent dans les groupes Euro American-213, Euro American-190, Euro American-246, Euro American-25, Euro American-163, Euro American-224 et Euro American-128. Ces 7 sous-lignées ont donc été réunies dans un groupe phylogénétique appelé *LAM-25/128/163/190/213/224/246*. Il faut rappeler que lors de précédentes études T5-RUS1 avait déjà été suspecté d'appartenir au groupe LAM (Gibson et al. 2008; Mokrousov et al. 2009; Abadia et al. 2010).

Haarlem-42/43/45 est observé partout en Europe, mais en le retrouve surtout en Europe du Sud, en Amérique du Sud et dans le Nord de l'Afrique, où il regroupe environ 20% des isolats modernes PGG2/3 (tableau 15B).

LAM-25/128/163/190/213/224/246 est observé sur les 3 continents Europe, Afrique et Amérique (avec 20% à 80% des isolats PGG2/3). Cependant, il est moins fréquent en Europe du Sud (7,5%), et en Asie de l'Ouest (4%), et absent en Afrique centrale et en Afrique de l'Ouest (tableau 15A). Fait intéressant : *LAM-25/128/163/190/213/224/246* regroupe presque la totalité des souches PGG2/3 en Afrique de l'Est (84.6%). Le fait que Euro American-213, qui est le premier à avoir divergé au sein de ce groupe, soit prévalent dans la région Nord de l'Afrique (13,43% des isolats PGG2/3 de la région), laisse supposer que cette région serait le lieu (ou proche du lieu) où *LAM-25/128/163/190/213/224/246* aurait pris naissance.

Les différentes sous-lignées de *LAM-25/128/163/190/213/224/246* ont chacune leur zone géographique de prédilection : en Afrique du Nord Euro American-213 est prévalent parmi les isolats PGG2/3 (13,43%), en Afrique de l'Est c'est Euro American-163 (50% des isolats PGG2/3), en Afrique du Sud c'est Euro American-246 (14,75%) et en Amérique du Sud c'est Euro American-190 (21,15% des isolats PGG2/3).

A

	LAM- 25/128/163/190/213/224/246	LAM- 190/213/246	LAM- 25/128/163/224
AFRI-E	84,62	13,64	86,36
AMER-C	58,33	14,29	85,71
AMER-S	43,27	54,81	45,19
AFRI-S	39,47	78,27	21,73
AFRI-N	37,31	68	32
CARI	34,07	10,48	89,52
EURO-S	26,25	41,18	58,82
EURO-W	23,27	39,72	60,28
AMER-N	22,41	54,19	45,81
EURO-N	20,56	54,9	45,1
EURO-E	7,46	80	20
ASIA-W	4	40	60
AFRI-M	0	0	0
AFRI-W	0	0	0

B

	Haarlem-42/43/45
EURO-S	26,64
AMER-S	23,56
AFRI-N	20,9
EURO-W	17,33
EURO-E	17,16
EURO-N	15,32
ASIA-W	13,6
CARI	12,64
AMER-N	11,45
AFRI-S	3,28
AFRI-W	2,28
AFRI-E	1,92
AFRI-M	0
AMER-C	0

Tableau 15. Distribution géographique de différents groupes phylogénétiques dans les diverses sous régions du monde. Pourcentage des souches appartenant à un groupe phylogénétique donné, dans une sous-région donnée, par rapports à l'ensemble des souches PGG2/3 répertoriées dans cette sous-région. A) Distribution des souches appartenant à *LAM-25/128/163/190/213/224/246* et à ses deux sous-groupes *LAM-190/213/246* et *LAM-25/128/163/224*. B) Distribution des souches appartenant à *Haarlem-42/43/45*.

3.2.4.3. Identification du groupe phylogénétique présentant la délétion RD(Rio)

La grande famille spoligotypique LAM (Brudey et al., 2006) a une sous-lignée qui se caractérise par la délétion RD(Rio) (Gibson et al., 2008).

L'arbre bayésien (figure 21) montre que la branche *LAM-25/128/163/190/213/224/246* se subdivise en deux sous-branches. La première que l'on va désigner « sous-branche A » (ou *LAM-190/213/246*) porte les sous-lignées : Euro American-213, Euro American-190 et Euro American-246. La seconde sous-branche que l'on va désigner « sous-branche B » (ou *LAM-25/128/163/224*) porte les

sous-lignées : Euro American-25, Euro American-163, Euro American-224 et Euro American-128. Il est intéressant de vérifier si les deux sous-branches de *LAM-25/128/163/190/213/224/246* peuvent aussi être caractérisées par la présence et l'absence de RD(Rio). Dans ce but, un échantillon de souches LAM déjà publié (Lazzarini et al., 2007), dans lequel certains isolats présentaient la délétion RD(Rio) alors que d'autres non, a été reclassifié selon la nouvelle classification basée sur les MIRU-VNTR (annexe 6).

3.2.4.3.1. Reclassification de profils MIRU-VNTR d'étude déjà publiée

Un échantillon de 190 souches publiées (Lazzarini et al., 2007) a été reclassifié. Cet échantillon contenait 93 isolats avec la région RD(Rio) délétée, 90 isolats avec une région RD(Rio) intacte et 7 isolats internationaux de référence sans la région RD(Rio). Sur les 100 souches présentant la délétion RD(Rio), 2% (n=2) n'ont pas été reconnues *LAM-25/128/163/190/213/224/246* d'après leur profils MIRU12, 95% (n=95) appartiennent à des sous-lignées de la « sous-branche B » et 3% (n=3) à des sous-lignées de la « sous-branche A ».

Sur les 90 souches présentant une région RD(Rio) intacte, 10% (n=9) ne furent pas reconnues *LAM-25/128/163/190/213/224/246* d'après leur profils MIRU12, 77,8% (n=70) appartiennent à des sous-lignées de la « sous-branche A » et 12,2% (n=11) à des sous-lignées de la « sous-branche B ».

On remarque donc que 86,4% (70/81) des souches ne présentant pas la délétion sont dans les lignées de la « sous-branche A » et que 96,9% (95/98) des isolats qui présentent la délétion RD(Rio) sont dans les lignées de la « sous-branche B »

3.2.4.3.2. Description du groupe phylogénétique RD(Rio)- dans le schéma de classification basé sur les 12 loci MIRU-VNTR

D'après les observations du paragraphe précédent, on peut conclure que les lignées de la « sous-branche A » se caractérisent par une région RD(Rio) intacte, alors que celles de la sous-branche B, n'ont pas la région RD(Rio). Ainsi l'ancêtre de *LAM-25/128/163/190/213/224/246* avait la région RD(Rio) lorsque que Euro American-213 et l'ancêtre de Euro American-190 et de Euro American-246 ont divergé. Toutes ces lignées forment donc le groupe *LAM-190/213/246* caractérisé par une région RD(Rio) intacte. Puis un des représentants de l'ancêtre de *LAM-*

retrouve dans la même famille MIRU-VNTR. On remarque qu'un bloc d'espaceurs absent chez l'un, est contigu à un bloc d'espaceurs absents chez l'autre. Par exemple dans le cas F, le profil 2 a perdu les espaceurs 10 à 19 alors que l'isolat 1 a perdu les espaceurs 20 à 24. Le premier est un LAM7-TUR et le second isolat du cas F est un T3-ETH. Ces deux profils se retrouvent classifiés dans la sous-lignée Euro American-40, Il faut ici rappeler que les 43 espaceurs révélés dans le spoligotypage₄₃ sont dispersés parmi les 94 différents espaceurs que compte le locus DR (van Embden et al. 2000). Les espaceurs 1 à 3, 4 à 7, 8 à 34, 35 à 36, 37 à 39 et 40 à 43 sont contigus dans le locus DR. Ainsi les pertes 10 à 19 du profile 2 et 20 à 24 du profile 1 sont réellement contiguës (cas F, tableau 16). Pour pouvoir expliquer ces discordances entre les deux classifications, il faut une vue plus claire du locus DR.

3.2.5.2. Implication probable de l'IS6110 dans le polymorphisme du locus DR

3.2.5.2.1. Mise en évidence d'une forte présence de d'IS6110 dans le locus DR

La technique de génotypage (Legrand et al. 2001) pour la détection des insertions de l'IS6110 dans le locus DR a été appliquée à 10 souches du Mozambique. Celle-ci a permis de détecter la présence d'IS6110 dans la partie du Locus DR où se trouvent les 86 premiers espaceurs. En appliquant la procédure de l'interprétation des résultats expliquée dans le chapitre « Matériels et Méthodes », il a été possible de faire une représentation du locus DR de la souche 1172 du Mozambique (figure 22B). Comme le montre la figure 22B, le locus DR de cette souche présente beaucoup de perte d'espaceurs : 4 à 11, 16, 32 à 33, 43 à 50, 54 à 61, 67 à 78, 80 à 84 et 86. L'interprétation des résultats nous a forcé à admettre qu'il y a des espaceurs dupliqués où un des doublons serait inversé (direction 3'-5'). Ce fait a déjà été rapporté (Caimi et al. 2001), (van der Zanden et al. 2002). Ce serait le cas pour les espaceurs 4 et 34 de la souche 1172 (figure 22C). Le locus DR de la souche 1172 présenterait 10 insertions d'IS6110 dans les DR suivants : DR2 (il s'agit du DR situé avant l'espaceur 2), DR4, DR12, DR17, DR23, DR27, DR29, DR31, DR34 and DR35. La présence d'un IS6110 dans le DR35 a déjà été reportée (Schurch et al. 2011).

3.2.5.2.2. Mise en évidence de délétions adjacentes aux zones d'insertion de l'élément transposable IS6110

Ces délétions qui sont situées à côté des insertions d'IS6110 ont déjà été observées dans une étude publiée (Lari et al. 2001).

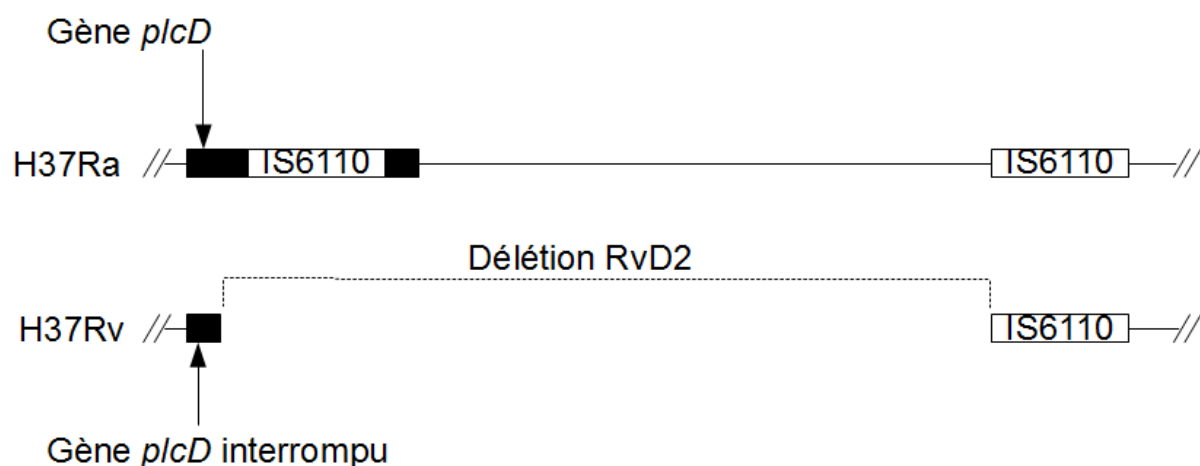


Figure 23. Polymorphisme dans la région RvD2 de *M. tuberculosis*. La représentation d'en haut illustre une région RvD2 intacte de H37Ra entourée par deux IS6110 dans H37Ra. La seconde représentation illustre une région RvD2 délétée chez une souche H37Rv.

Dans cette étude, une souche H37Ra présente dans son génome deux IS6110 qui encadrent la région RvD2 (figure 23). Un des deux IS6110 se retrouve dans le gène *plcD* qui code pour une phospholipase C. La séquence génomique située entre les deux IS6110 est délétée chez un H37Rv, et celui-ci ne contient qu'une seule IS6110 adjacente à la région délétée. Selon Sampson et al., (2003), une recombinaison génomique entre deux IS6110 conduit à la perte de la région située entre les 2 séquences d'insertion. Dans le même esprit, Ho et al., en 2000, ont montré que les régions de 20000 pb qui encadrent le point d'insertion d'un IS6110, sont très variables et sujettes aux délétions (Ho et al. 2000). De même la région génomique allant de Rv1754 à Rv1765, (qui correspond à RD152), est sujette à des délétions (Sampson et al. 2003). Il est important de remarquer que certains LSP sont situés à côté d'un IS6110, C'est le cas pour RD152, RD207, RD5, RD11, RD14, MiD2 (Brodin et al. 2002; Rao et al. 2005; Tsolaki et al. 2005) (tableau 17).

Pour déterminer si l'IS6110 est impliquée dans des délétions génomiques, l'IS6110AD-typing a été appliqué sur 10 souches du Mozambique et un H37Rv (contrôle positif) (tableau 18). Certaines régions n'ont pas été amplifiées, ce qui prouve qu'il y a des régions délétées à côté de l'IS6110, Ce qui est surprenant, c'est

que des délétions furent observées même chez l'H37Rv génotypé, alors que ces régions sont présentes dans le génome du H37Rv récupéré sur le serveur NCBI.

Nom des loci portant les 2 transposases d'IS6110	Région de différence	Nom des loci des gène(s) délétés
Rv1756c-Rv1757c	152	Rv1754c-Rv1765c (Tsolaki et al., 2005)
Rv2814c-Rv2815c	207	Rv2814c-Rv2819c and a part of Rv2820c (Tsolaki et al., 2005)
Rv2354-Rv2355	5	Rv2346c-Rv2353c (Rao et al., 2005)
Rv2648-Rv2649	11	Rv2645-Rv2659c (Rao et al., 2005)
Rv1763-Rv1764	14	Rv1765c-Rv1773c (Rao et al., 2005)
Rv3186-Rv3187	MI2	Rv3188-Rv3189 (Brodin et al., 2002)

Tableau 17. Description de certaines régions de différence (RD) occupant une position adjacente à un IS6110, La première colonne donne les noms des loci des deux transposases portées par un IS6110 donné. La seconde colonne liste les RD adjacents à un IS6110 donné. La troisième colonne liste les noms des loci concernés par une délétion donnée.

Ceci est certainement dû aux nombreux repiquages permettant la conservation de la souche H37Rv. En plus le spoligotype43 de cet H37Rv génotypé a subi une délétion additionnelle (espaceur 15). Cependant le pattern MIRU12 de cette souche H37Rv reste inchangé.

IS6110 n°	1		2		3		4		5		6		7		8		9		10		12		13		14		15		16	
r = région en 5'/l = région en 3' par rapport à un IS6110	r	l	r	l	r	l	r	l	r	l	r	l	r	l	r	l	r	l	r	l	r	l	r	l	r	l	r	l	r	l
H37Rv	□	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1145	□	■	□	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1523	□	■	□	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1264	□	■	□	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1218	□	■	□	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1369	□	■	□	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1172	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■
1368	■	■	■	■	□	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
1371	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■
1360	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■
1456	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■	□	■

Tableau 18. Résultats de l'IS6110AD-typing appliqué sur 10 souches du Mozambique. Un carré plein symbolise une région intacte, un carré vide symbolise une région délétée. Une astérisque est ajoutée lorsque la région présente environ 50 bp en moins que la taille attendue.

L'application de l'IS6110AD-typing sur les 10 souches du Mozambique montre qu'il y a des délétions adjacentes au point d'insertion des IS6110, Cependant, on sait que le locus DR est un *ipl* (IS6110 preferential locus), (Fang and Forbes 1997). Par exemple, la souche 1172 compte visiblement dix IS6110 dans son locus DR (figure 22C).

Plusieurs mécanismes sont suggérés pour expliquer la perte d'espaceurs dans les loci CRISPR : recombinaison homologue entre DRs (Fang et al. 1998), des

glissements de la polymérase lors de la réplication de l'ADN (Jansen et al. 2002) et les recombinaisons homologues entre *IS6110* (Sampson et al. 2003). Dans une étude de Schürch et al., (2011), des spoligotypes différents ont été observés chez des souches liées épidémiologiquement. Selon eux, la perte d'un bloc d'espaceurs était due à une recombinaison entre DRs qui a été favorisée par la formation d'un élément secondaire impliquant deux *IS6110* insérées dans des directions opposées. Qu'importe le mécanisme, le fait que les *IS6110* soient observées à côté des pertes d'espaceurs (Figure 22C), conforte l'idée qu'elles soient impliquées dans les délétions survenant dans le locus DR.

3.2.5.3. Comment expliquer les discordances entre la classification spoligotypique et celle basée sur les MIRU-VNTR

Revenons au cas F du tableau 16, l'ancêtre que les deux profils 1 et 2 ont en commun, devait avoir tous les espaceurs de 10 à 24 (numéro faisant référence à la position génomique) et en plus un *IS6110* dans le DR20 (DR situé entre les espaceurs 19 et 20). Cet ancêtre aurait tantôt connu une délétion adjacente en amont de l'*IS6110* et tantôt en aval, pour mener au profil 2 (espaceurs 10 à 19 absents) et au profil 1 (espaceurs 20 à 24 absents). Cet événement aurait mené à deux patterns spoligotypiques distincts classifiés abusivement dans 2 sous-lignées spoligotypiques distinctes. Le fait que la nouvelle classification MIRU-VNTR mettent ces deux patterns dans la même sous-famille ne serait donc pas erroné, mais au contraire il nous amène à une autre réalité : des souches présentant des spoligotypes soit disant divergents peuvent être très proches phylogénétiquement.

Dans le cas C du tableau 16, deux profils qui diffèrent sur l'état d'un espaceur sont dans deux sous-lignées spoligotypiques différentes, alors qu'ils se retrouvent dans la même sous-famille MIRU-VNTR. L'insertion d'*IS6110* dans un DR peut être asymétrique, ce qui masque la présence d'un des deux espaceurs adjacents à cause d'une absence d'amplification lors de la PCR.

Les souches appartenant aux lignées spoligotypiques modernes Haarlem, LAM, X, S, T présentent beaucoup d'*IS6110* par opposition à la lignée ancestrale EAI qui présente peu ou pas d'*IS6110*. Ce nombre élevé de copies d'*IS6110* chez les PGG2/3 causerait beaucoup de variation dans leur locus DR, rendant difficile l'étude

de leur évolution à partir de leur profil spoligotypique et l'identification, sans confusion, des diverses lignées spoligotypiques parmi les souches PGG2/3.

4. Conclusion générale et perspectives

4.1. Intérêt global d'une méthodologie de classification basée sur les minisatellites VNTR

La classification MIRU-VNTR présente 2 intérêts majeurs:

- Elle apporte un nouvel angle de vue à la classification des souches chez MTC. Deux angles avaient déjà été posés : la classification spoligotypique (Brudey et al., 2006 ; Demay et al., 2012) et la classification LSP (Gagneux et al., 2007). Elle a permis de soulever les erreurs dans la classification spoligotypique des souches modernes PGG2/3, erreurs que l'on avait déjà suspectées avec les SNPs (Abadia et al., 2010). Elle permet aussi d'avoir une classification plus détaillée, (subdivision plus fine), que celle basée sur les LSPs.
- Cette classification est d'autant plus importante pour les souches modernes PGG2/3 qui présentent beaucoup d'*S6110*, et pour lesquelles l'application de la classification spoligotypique est plus délicate.

4.1.1. Intégration des lignées génotypiques de la classification MLVA au modèle VNTR-TBbayes

À la lumière des résultats obtenus dans le cadre de cette thèse, la version de VNTR-TBbayes mise en ligne contient 14 variables :

- 12 variables relatives aux 12 loci MIRU-VNTR. Ces variables ont pour attributs un ensemble de valeurs correspondant aux différentes valeurs de répétitions observées dans l'échantillon utilisé pour faire l'apprentissage de structure,
- une variable relative à la région d'origine prenant pour états 20 régions différentes qui couvrent le monde,
- une variable relative à la lignée génotypique. Celle-ci à 41 états correspondant aux 41 lignées nouvellement définies et basées sur le marqueur MIRU-VNTR.

Il semblait plus cohérent d'écarter les profils de spoligotypage du modèle puisque ceux-ci ont déjà un système de classification qui leur est propre (Brudey et al., 2006 ; Demay et al., 2012). Cependant, en ce qui concerne la classification MLVA, les paramètres (moyenne et écart-type pour chaque famille génotypique) de la méthode

CZS n'ont pas été publiés. De plus aucun outil n'a pas encore mis à disposition des scientifiques du domaine pour qu'ils puissent intégrer cette méthode de classification à leurs études. Il a donc été nécessaire de mettre en place VNTR-TBbayes pour pallier à ce manquement.

Les paramètres (moyenne et écart-type pour chaque famille génotypique) de la méthode CZS sont tributaires de l'échantillon. Si ce dernier n'est pas exhaustif (ou représentatif) au sein d'une lignée, cela crée un biais qui affecte indubitablement la classification. Le modèle bayésien, en se basant sur les relations d'influence existant entre les différentes variables (régions, loci MIRU, lignées), peut corriger ce biais. De plus la méthode CZS assigne une lignée à un pattern mais ne dévoile pas le degré de confiance que l'on peut avoir dans cette assignation. Un système probabilisé comme VNTR-TBbayes attribue à un pattern une liste de lignées probables, et pour chacune d'elles un degré de certitude est donné.

L'ensemble des 176 isolats de MIRU-VNTR^{plus} ont été reclassifiés selon la méthode CZS et selon VNTR-TBbayes (Annexe 5). Dans quelques cas (n=12), la lignée la plus probable prédite par VNTR-TBbayes ne correspond pas à celle trouvée par la méthode CZS. La souche 9550/00 qui est classifiée à la fois West African lineage II et *M. bovis* par la méthode CZS, est classifiée West African II par VNTR-TBbayes avec une probabilité de 0,98 (Annexe 5). La classification LSP met également cette souche parmi West African II.

L'annexe 7 montre la distribution des lignées assignées par VNTR-TBbayes à l'intérieur des lignées CZS. Par exemple, parmi les souches classifiées Indo-Oceanic-57 par la méthode CZS (= cluster 57), 99% et 1% sont respectivement prédites Indo-Oceanic-57 et Indo-Oceanic-59 par VNTR-Tbbayes. Cette annexe prouve que dans certains cas (minoritaires), la lignée la plus probable prédite par VNTR-TBbayes, ne correspond pas à celle assignée par la méthode CZS. Par exemple, parmi les patterns MIRU-VNTR prédits East Asian-93 par la méthode CZS, 78,69% ont également été prédits East Asian-93 par VNTR-TBbayes, 14,75% ont été prédits East-African Indian-68, les autres patterns ont été classifiés East Asian-17 (4,92%) et East Asian-16 (1,64%). La proportion assez importante de patterns MIRU-VNTR East Asian-93 qui ont été reclassifiées East-African Indian-68 par VNTR-TBbayes, sous-entend une forte proximité phylogénétique entre ces deux lignées.

4.1.2. Réévaluation de la classification spoligotypique sur la base de la classification MIRU-VNTR

Le tableau 19 montre la distribution des lignées MIRU-VNTR à l'intérieur des lignées spoligotypiques. Par exemple, parmi les souches classifiées AFRI1 (= Cluster AFRI1), 63,64% sont prédites West African II par VNTR-Tbbayes. Ce tableau met en évidence la convergence que masquent les lignées spoligotypiques. Par exemple, parmi le grand groupe spoligotypique LAM certains sous-groupes ne seraient pas des LAM. C'est le cas de LAM7-TUR (Abadia et al., 2010) et de LAM10-CAM (Niobe-Eyangoh et al., 2004). À l'inverse, d'autres sous-lignées spoligotypiques reconnues pour être des T seraient des LAM : c'est le cas de H37Rv, T5-RUS1 (Mokrousov et al., 2009) et T5-Madrid2. Les règles spoligotypiques qui permettent l'identification de vraies souches LAM sont celles définissant les lignées : LAM1, LAM2, LAM3, LAM4, LAM5, LAM6, LAM8, LAM9, LAM11-ZWE, LAM12-Madrid1, H37Rv, T5-RUS1, T5-Madrid2 et également T-Tuscany. Cependant, ces sous-lignées spoligotypiques ne regroupent pas uniquement des isolats appartenant au groupe LAM-25/128/163/190/213/224/246. Par exemple, 91,67% des souches LAM1 appartiennent au groupe LAM-25/128/163/190/213/224/246, 60,71% des souches LAM5 sont prédites LAM-25/128/163/190/213/224/246 par VNTR-TBbayes etc (voir Tableau 19). Le tableau 19 permet également d'évaluer la proportion de souches présentant la délétion RD(Rio) dans les sous-lignées spoligotypiques LAM. Par exemple, 85% des souches LAM1 sont des LAM-25/128/163/224, groupe génotypique qui se caractérise par la perte de la région RD(Rio).

Vu que le groupe spoligotypique Haarlem correspond aux lignées Euro American-42, Euro American-43 et Euro American-45, on peut conclure que la lignée spoligotypique H4 n'est pas un Haarlem, puisqu'elle ne regroupe aucune souches de ces 3 lignées MIRU-VNTR.

Les souches *M. microti* et *M. pinnipedii* sont majoritairement des West African II selon VNTR-TBbayes.

Cluster AFRI1		Cluster EAI2 *	
West African II	63,64	Euro American-15	100
West African I	22,73	Cluster EAI2-Manilla	
Euro American-121	4,55	Indo-Oceanic-56	100
Euro American-33	4,55	Cluster EAI2-nonthaburi	
Indo-Oceanic-56	4,55	Indo-Oceanic-56	100
Cluster AFRI3		Cluster EAI3-IND	
West African I	100	Indo-Oceanic-69	63,46
Cluster AFRI		Indo-Oceanic-57	25,00
West African II	43,75	Indo-Oceanic-56	3,85
West African I	37,50	Euro American-15	1,92
East-African Indian-68	6,25	Euro American-43	1,92
Euro American-34	6,25	Euro American-46	1,92
Euro American-8	6,25	Indo-Oceanic-64	1,92
Cluster Beijing		Cluster EAI4-VNM	
East Asian-17	33,47	Indo-Oceanic-59	92,86
East Asian-16	31,17	Indo-Oceanic-57	7,14
East Asian-99	9,08	Cluster EAI5	
East Asian-86	6,50	Indo-Oceanic-57	38,37
East Asian-93	6,23	Indo-Oceanic-59	25,58
East Asian-101	5,01	Indo-Oceanic-64	16,28
East Asian-83	5,01	Indo-Oceanic-69	10,47
Euro American-116	1,76	Indo-Oceanic-56	6,98
Euro American-121	0,27	Euro American-121	1,16
Euro American-15	0,27	Euro American-7	1,16
Euro American-34	0,27	Cluster EAI6-BGD1	
Euro American-40	0,27	Indo-Oceanic-57	59,09
Euro American-8	0,27	Indo-Oceanic-56	18,18
Euro American-128	0,14	Indo-Oceanic-59	13,64
Euro American-212	0,14	Indo-Oceanic-64	9,09
Euro American-33	0,14	Cluster EAI8-MDG	
Cluster BOV 1		Indo-Oceanic-57	62,50
M. bovis	96,67	Indo-Oceanic-64	37,50
Euro American-33	1,67	Cluster H1	
Euro American-42	1,67	Euro American-43	29,45
Cluster BOV 2		Euro American-42	23,29
M. bovis	94,44	Euro American-45	18,49
West African II	5,56	Euro American-46	8,22
Cluster BOV 3		Euro American-121	3,42
M. bovis	100	Euro American-112	2,05
Cluster BOV 4-CAPRAE		Euro American-213	2,05
West African II	99,17	Euro American-34	2,05
Euro American-15	0,83	Euro American-7	2,05
Cluster BOV		Euro American-190	1,37
M. bovis	96,77	Euro American-212	1,37
Euro American-7	3,23	Euro American-33	1,37
Cluster CANETTII		East Asian-16	0,68
M. canettii	100	Euro American-116	0,68
Cluster CAS1-Delhi		Euro American-12	0,68
East-African Indian-68	69,64	Euro American-125	0,68
East Asian-101	10,71	Euro American-15	0,68
East Asian-93	8,93	Euro American-246	0,68
East Asian-17	5,36	Euro American-8	0,68
East-African Indian-261	1,79	Cluster H2	
East Asian-86	1,79	Euro American-42	55,56
Euro American-15	1,79	Euro American-43	33,33
Cluster CAS1-Kili		Euro American-112	11,11
East-African Indian-261	100		
Cluster CAS2			
East-African Indian-68	66,67		
East Asian-93	33,33		
Cluster CAS			
East-African Indian-68	57,58		
Euro American-34	15,15		
Euro American-33	9,09		
East Asian-101	6,06		
East-African Indian-261	3,03		
East Asian-83	3,03		
East Asian-93	3,03		
West African I	3,03		
Cluster EAI1-SOM			
Indo-Oceanic-64	70,83		
Indo-Oceanic-59	16,67		
Indo-Oceanic-57	12,50		

Cluster H3		Cluster LAM6	
Euro American-42	26,96	Euro American-190	85,19
Euro American-45	21,84	Euro American-246	11,11
Euro American-46	7,85	Euro American-224	3,70
Euro American-40	7,51	Cluster LAM7-TUR	
Euro American-43	6,83	Euro American-40	98,11
Euro American-33	6,48	Euro American-224	1,89
Euro American-7	5,46	Cluster LAM8	
Euro American-213	4,44	Euro American-25	100
Euro American-15	2,39	Cluster LAM9	
Euro American-8	2,05	Euro American-190	24,38
East Asian-101	1,71	Euro American-246	20,63
Euro American-116	1,37	Euro American-25	13,75
Euro American-125	1,02	Euro American-213	10
Euro American-34	1,02	Euro American-34	8,75
East Asian-17	0,68	Euro American-128	6,25
Euro American-112	0,68	Euro American-163	5
East-African Indian-261	0,34	Euro American-224	4,38
Euro American-12	0,34	Euro American-15	1,88
Euro American-121	0,34	Euro American-33	1,25
Euro American-224	0,34	Euro American-8	1,25
Euro American-25	0,34	Euro American-112	0,63
Cluster H4		Euro American-121	0,63
Euro American-33	75	Euro American-46	0,63
Euro American-7	25	Euro American-7	0,63
Cluster H37Rv		Cluster LAM10-CAM	
Euro American-128	100	Euro American-12	86,96
Cluster LAM1		Euro American-34	2,90
Euro American-128	58,33	Euro American-112	1,45
Euro American-25	25	Euro American-121	1,45
Euro American-246	5	Euro American-128	1,45
East Asian-101	3,33	Euro American-212	1,45
Euro American-12	1,67	Euro American-224	1,45
Euro American-15	1,67	Euro American-46	1,45
Euro American-163	1,67	Euro American-7	1,45
Euro American-190	1,67	Cluster LAM11-ZWE	
Euro American-8	1,67	Euro American-163	96,67
Cluster LAM2		Euro American-212	3,33
Euro American-25	66,67	Cluster LAM12-Madrid1	
Euro American-33	8,33	Euro American-213	90
Euro American-34	8,33	Euro American-25	10
East-African Indian-68	4,17	Cluster MANU2	
East Asian-16	4,17	Euro American-116	20
Euro American-128	4,17	Euro American-128	20
Euro American-224	4,17	Euro American-40	20
Cluster LAM3		Euro American-46	20
Euro American-213	58,68	Indo-Oceanic-64	20
Euro American-25	12,57	Cluster MANU3	
Euro American-224	7,78	Euro American-34	100
Euro American-212	4,79	Cluster microti	
Euro American-190	2,99	West African II	100
Euro American-128	2,40	Cluster PINI1	
Euro American-246	2,40	West African II	66,67
Euro American-121	1,80	Indo-Oceanic-56	33,33
Euro American-125	1,20	Cluster PINI2	
Euro American-33	1,20	West African II	100
Euro American-34	1,20	Cluster S	
Euro American-8	1,20	Euro American-212	71
East Asian-101	0,60	Euro American-15	8
Euro American-116	0,60	Euro American-116	3
Euro American-43	0,60	Euro American-121	3
Cluster LAM4		Euro American-42	3
Euro American-246	38,10	East Asian-101	2
Euro American-163	23,81	Euro American-125	2
Euro American-190	19,05	Euro American-128	2
Euro American-25	14,29	Euro American-43	2
Euro American-224	4,76	Euro American-213	1
Cluster LAM5		Euro American-33	1
Euro American-25	53,57	Euro American-7	1
Euro American-121	17,86	Indo-Oceanic-57	1
Euro American-34	14,29	Cluster T-tuscany	
East Asian-101	3,57	Euro American-190	50
Euro American-213	3,57	Euro American-33	50
Euro American-246	3,57		
Euro American-7	3,57		

Cluster T1	
Euro American-15	18,62
Euro American-8	17,76
Euro American-224	7,41
Euro American-121	7,07
Euro American-112	6,55
Euro American-190	5
Euro American-7	4,48
Euro American-40	4,31
Euro American-116	4,14
Euro American-34	3,97
Euro American-125	3,45
Euro American-33	3,10
Euro American-212	2,24
East Asian-101	1,90
Euro American-128	1,90
Euro American-45	1,55
Euro American-46	1,55
Euro American-12	1,21
Euro American-213	1,03
Euro American-25	0,86
Euro American-42	0,52
East Asian-16	0,34
East Asian-86	0,34
Euro American-246	0,17
Euro American-43	0,17
Indo-Oceanic-56	0,17
Indo-Oceanic-64	0,17
Cluster T1-RUS2	
Euro American-8	76,92
East Asian-83	15,38
Euro American-7	7,69
Cluster T2	
Euro American-15	19,54
Euro American-116	17,24
Euro American-8	10,34
Euro American-34	6,90
Euro American-121	5,75
Euro American-125	5,75
Euro American-190	5,75
East Asian-86	3,45
Euro American-112	3,45
Euro American-40	3,45
Euro American-7	3,45
Euro American-12	2,30
Euro American-128	2,30
Euro American-224	2,30
East Asian-101	1,15
East Asian-16	1,15
East Asian-17	1,15
Euro American-212	1,15
Euro American-246	1,15
Euro American-33	1,15
Euro American-46	1,15
Cluster T2-uganda	
Euro American-121	37,50
Euro American-12	31,25
Euro American-112	12,5
East Asian-101	6,25
Euro American-15	6,25
Euro American-34	6,25
Cluster T3	
Euro American-15	36,67
Euro American-112	23,33
Euro American-8	10
Euro American-121	6,67
Euro American-116	3,33
Euro American-125	3,33
Euro American-212	3,33
Euro American-224	3,33
Euro American-40	3,33
Euro American-7	3,33
Indo-Oceanic-69	3,33
Cluster T3-ETH	
Euro American-40	100
Cluster T3-OSA	
Euro American-40	100
Cluster T4	
Euro American-112	42,86
Euro American-15	14,29
Euro American-34	14,29
Euro American-7	14,29
Euro American-8	14,29
Cluster T4-CEU1	
Euro American-34	84,21
Euro American-121	14,04
Euro American-15	1,75
Cluster T5	
Euro American-15	45,45
Euro American-224	31,82
Euro American-213	9,09
Euro American-121	4,55
Euro American-212	4,55
Euro American-8	4,55
Cluster T5-Madrid2	
Euro American-224	100
Cluster T5-RUS1	
Euro American-190	100
Cluster X1	
Euro American-34	41,24
Euro American-15	12,43
Euro American-45	7,34
Euro American-112	6,78
Euro American-121	6,21
Euro American-125	5,65
Euro American-33	5,65
Euro American-46	5,65
Euro American-190	2,82
Euro American-116	1,69
Euro American-213	1,69
East-African Indian-68	0,56
Euro American-163	0,56
Euro American-246	0,56
Euro American-25	0,56
Euro American-40	0,56
Cluster X2	
Euro American-33	50,96
Euro American-34	14,42
Euro American-15	13,46
Euro American-121	7,21
Euro American-116	6,25
Euro American-213	3,37
Euro American-7	0,96
Indo-Oceanic-69	0,96
Euro American-12	0,48
Euro American-45	0,48
Euro American-46	0,48
Indo-Oceanic-56	0,48
Indo-Oceanic-64	0,48
Cluster X3	
Euro American-34	29,17
Euro American-7	19,17
Euro American-33	16,67
Euro American-121	10,83
Euro American-15	9,17
Euro American-125	5,83
Euro American-46	5
Euro American-45	1,67
Euro American-8	1,67
Euro American-213	0,83
Cluster ZERO*	
Indo-Oceanic-59	100

Tableau 19. Distribution des lignées prédites par VNTR-TBbayes à l'intérieur des lignées décrites par Brudey (Brudey et al, 2006). Un astérisque (*) suit les clusters qui ne contiennent qu'une seule souche. Ce tableau fut réalisé à partir de 4313 souches de SITVITWEB, décrites avec des résultats de spoligotypage et de MIRU-VNTR.

4.1.3. Un exemple d'évolution convergente dans le locus DR de *Mycobacterium tuberculosis* (Fenner et al., 2011).

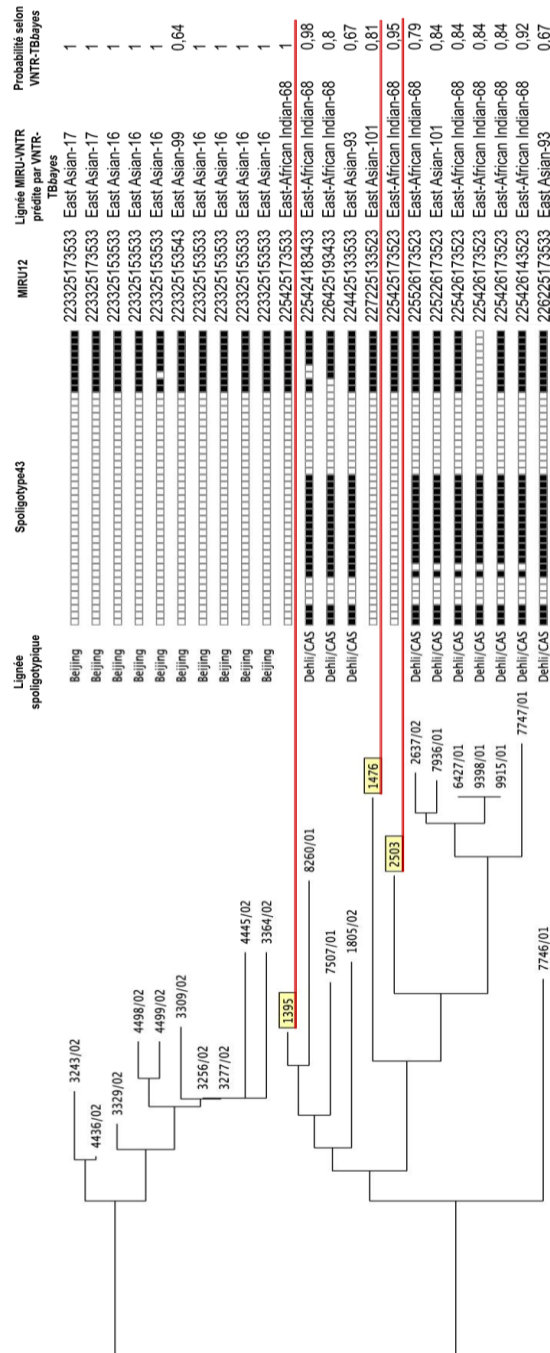


Figure 24. Arbre Neighbor-Joining réalisé avec les résultats de typage de 24 loci MIRU-VNTR de 23 isolats (Fenner et al., 2011). Cette figure tirée de l'article de Fenner et al. en 2011 a été modifiée. Le MIRU12 (tiré du MIRU24 publié) correspondant à chaque isolat a été ajouté et la lignée MLVA fut prédite par VNTR-TBbays.

Dans une étude de Fenner et al. en 2011, trois isolats présentent un pattern spoligotypique identique à la signature Beijing, alors que les résultats de deux autres

marqueurs (SNP, LSP) sont en faveur d'une assignation CAS. L'arbre Neighbor-Joining réalisé avec les résultats de typage de 24 loci MIRU-VNTR place ces trois souches (soulignées par un trait rouge dans la figure 24) sur la branche portant les isolats CAS. En effet, le SNP spécifique aux isolats Beijing s'est révélé négatif, et celui spécifique aux souches CAS est positif. De plus, ces trois isolats présentent la région RD105 (dont la délétion est spécifique aux souches Beijing), alors que la région RD750 était absente (ce qui est une caractéristique des souches CAS). La reclassification de ces trois souches par VNTR-TBbayes grâce à leur profil MIRU12, révèle que deux d'entre elles appartiendraient à la lignée East-African Indian-68 avec une probabilité de 1 et de 0,95 alors que la troisième serait un East Asian-101 avec une probabilité de 0,81. La branche de l'arbre de la figure 24 qui porte les profils CAS, se compose de souches East-African Indian-68, East-African Indian-93 et East-African Indian-101. On peut donc conclure que les trois lignées MLVA sont proches.

4.2. Le passage à une classification basée sur les 24 loci MIRU-VNTR : est-ce un impératif ?

Pour répondre à cette question deux arbres phylogénétiques ont été construits avec 95 isolats du Mozambique. Un arbre fut construit avec 24 loci MIRU-VNTR (figure 25B) et l'autre avec les 12 loci classiques (figure 25A). Les deux arbres font une séparation entre les deux lignées Indo-Oceanic (en rouge) et Euro American (en bleu). On retrouve de plus les mêmes sous-clusters au sein de ces deux lignées. L'arbre B avec les 24 loci MIRU-VNTR est plus ramifié, cependant, les 12 loci MIRU-VNTR sont suffisants pour la définition des lignées phylogénétiques puisque cela implique les souches MTC dans leur globalité à travers le monde. Le fort pouvoir discriminant des 24 loci trouve plus sa place dans des études visant à caractériser des souches proches s'étalant sur des régions géographiques réduites.

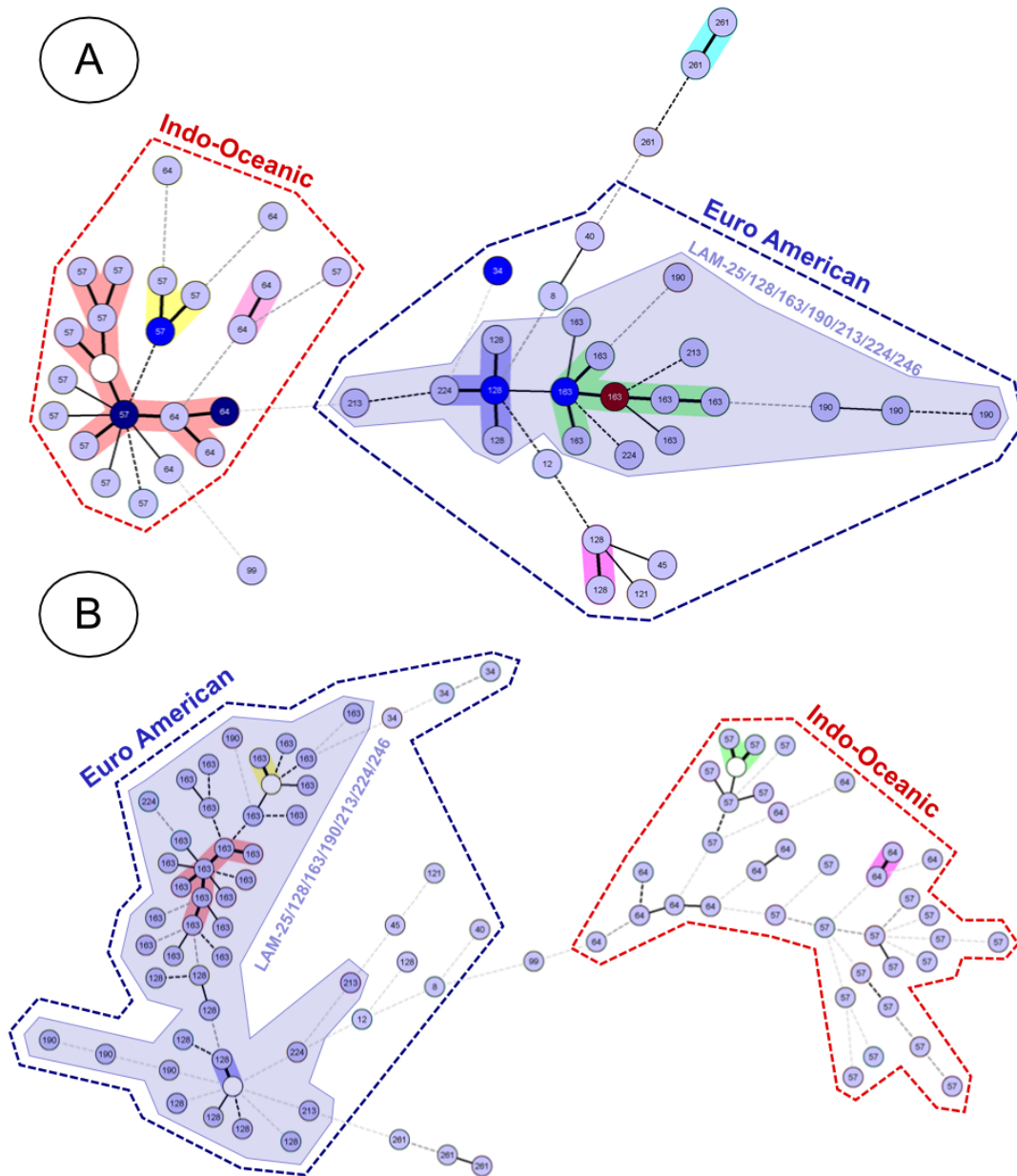


Figure 25. Arbre MST construit avec 95 souches du Mozambique. A) Arbre construit avec les 12 loci MIRU-VNTR. B) Arbre construit avec les 24 loci MIRU-VNTR.

4.3. Illustration de la géospécificité des lignées basées sur les MIRU-VNTR, relation avec l'histoire du peuplement du monde (Out of Africa)

Comme l'a introduit Alfred Russel Wallace, chaque espèce a une dispersion géographique propre, façonnée par plusieurs facteurs. La biogéographie des groupes phylogénétiques MRU-VNTR, nouvellement décrite, est indubitablement liée à l'histoire des peuplements. Pour vérifier cela, la distribution des lignées et sous-lignées MIRU-VNTR est représentée par la carte de la figure 26. Seules les lignées

Spécificité géographique de la lignée MIRU-VNTR Indo-Oceanic pour les régions bordant l'océan indien. Indo-Oceanic, qui, selon l'arbre bayésien de la figure 21, serait la plus ancienne des lignées *M. tuberculosis* (Wirth et al., 2008), serait originaire de la côte Est africaine. En effet, Indo-Oceanic-57 et Indo-Oceanic-64 seraient les premières sous-lignées Indo-Oceanic à émerger (plus proches du nœud central MIT57). D'ailleurs Indo-Oceanic-57 et Indo-Oceanic-64 sont prévalentes en Afrique de l'Est avec environ respectivement 60% et 40% des isolats PGG1 de *M. tuberculosis*. Les prochaines sous-lignées à apparaître seraient Indo-Oceanic-69 et Indo-Oceanic-56 qui partageraient un ancêtre proche. Comme le montre la figure 26, Indo-Oceanic-69 est prévalent en Asie du Sud (environ 60% des isolats Indo-Oceanic), et Indo-Oceanic-56 prédomine en Asie du Sud-Est (environ 60% des isolats Indo-Oceanic). Indo-Oceanic-69 aurait divergé avant Indo-Oceanic-56 d'après l'arbre bayésien (Figure 21). Ces spécificités géographiques suggèrent une migration en partance de l'Afrique de Est qui est arrivée dans les îles du Pacifique en passant par l'Arabie Saoudite et l'Inde. En effet au Moyen-Orient, presque toutes les souches Indo-Oceanic se retrouvent en Arabie Saoudite, alors que l'Inde concentre presque toutes celles de la sous-lignée Indo-Oceanic-69 de l'Asie du Sud. Ce trajet migratoire a déjà été décrit dans la littérature par des études réalisées sur l'haplogroupe mitochondrial M. Les lieux de prévalence de l'haplogroupe M, un des deux descendant de l'haplogroupe L3 (localisé en Afrique de l'Est), a permis de conclure qu'une des premières migrations humaines serait partie de la corne de l'Afrique pour rejoindre la côte arabe et persique (Underhill et Kivisild, 2007), (Renfrew, 2010). Puis un groupe d'individus aurait atteint l'Inde et la Thaïlande, l'Indonésie puis l'Australie (Renfrew 2010 ; Majumder 2010 ; Kayser 2010).

Le continent asiatique : lieu d'expansion de la lignée « East Asian and African Indian ». D'après l'arbre bayésien (figure 22) East Asian and African Indian (EAAI) occupe une position intermédiaire entre Indo-Oceanic et Euro American. Les sous-lignées de EAAI, qui prédomine en Asie, présentent aussi des spécificités géographiques propres. Il y a une grande proportion de East Asian-17 en Asie de l'Est (43,7% des isolats EAAI). En Asie centrale et du Nord, c'est East Asian-16 qui prédomine (avec respectivement 74.68% et 68.4% des isolats EAAI). Une étude de Mokrousov et al., en 2005 a démontré la présence de deux types de Beijing en Asie. Le premier se caractérise par une région NTF sans insertion d'IS6110, Selon lui,

celui-ci aurait émergé en Asie du centre au paléolithique supérieur et serait originaire du Moyen-Orient (Mokrousov et al. 2005). Les descendants de ce peuple, qui portait l'haplogroupe NRY K-M9, se seraient déplacés vers le Nord-Est en direction de la Sibérie et vers le Sud-Est en direction du Nord de la Chine (Mokrousov et al. 2005). Ce groupe Beijing qui prévaut dans le centre et le Nord de l'Asie pourrait correspondre à East Asian-16. Le second groupe de Beijing décrit par Mokrousov et al., (2005) présente un IS6110 dans la région NTF et aurait émergé au néolithique parmi les fermiers Proto-Sino-Tibetan (Haplogroup O-M214/M122) en Asie de l'Est. Il se serait ensuite étendu à l'ensemble de l'Asie de l'Est. La forte prédominance de East Asia-17 dans cette partie du monde laisse penser qu'il s'agirait du groupe de Beijing au locus NTF présentant une insertion d'IS6110,

Des études supplémentaires sont nécessaires pour valider ou rejeter cette hypothèse. Il serait tout aussi intéressant de vérifier où se trouvent les 3 LSPs RD181, RD142, RD150 caractéristiques des sous-groupes Beijing (Tsolaki et al., 2005) parmi les sous-lignées MIRU-VNTR de EAAI.

La lignée Euro American s'est probablement d'abord dispersée en Europe grâce à plusieurs vagues migratoires en provenance du Moyen-Orient. La lignée Euro American est la dernière à émerger parmi les *M. tuberculosis* d'après l'arbre bayésien (figure 21) si l'on se base sur le fait que Indo-Oceanic soit la lignée la plus ancestrale (Brosch et al. 2002). D'après l'arbre bayésien (Figure 21), Euro American-40 serait le plus ancien sous-groupe de Euro American. Euro American-40 est prévalente en Asie de l'Ouest avec 37,6% des sous PGG2/3 de la région. Cela suggère que Euro American serait apparu dans cette zone géographique du Moyen-Orient.

En général, la distribution des sous-lignées PGG2/3 dans les diverses sous-régions, est très hétérogène. Ce n'est pas le cas pour l'Afrique centrale et de l'Ouest où Euro American-12 compte 100% et 90,48% des isolats PGG2/3. Pour mieux comprendre cette forte prévalence de Euro American-12, la trajectoire de l'haplogroupe R1b (chromosome Y) peut être utile. En effet, c'est le seul haplogroupe qui soutient l'hypothèse d'une migration rétrograde d'Asie à l'Afrique de l'Ouest. L'haplogroupe R1b est le plus fréquent en Europe de l'Ouest, dans une partie du centre de l'Eurasie, et dans une partie de l'Afrique subsaharienne (autour du Tchad et du Cameroun) ([http://en.wikipedia.org/wiki/Haplogroup_R1b_\(Y-DNA\)#Origin_and_dispersal](http://en.wikipedia.org/wiki/Haplogroup_R1b_(Y-DNA)#Origin_and_dispersal)).

Le point d'origine de cet haplogroupe fut placé en Asie de l'Ouest (Myres et al. 2011). Cruciani et al. (2010) ont démontré par l'étude des haplogroupes du chromosome Y, un retour migratoire d'Asie vers l'Afrique au néolithique. Pour tester cette hypothèse, 154 souches publiées du Nord de l'Iran (Asgharzadeh et al. 2011), ont été reclassifiées en utilisant les résultats de typage des 12 loci MIRU-VNTR. Les trois premières lignées qui dominent dans cette région Euro American-212 (avec 22,8% des isolats), *M. bovis* (21,43%) et Euro American-121 (11,69%). On sait que Euro American-121 tout comme Euro American-12 comprend des souches d'origine africaine (annexe 5). En effet la reclassification des souches MIRU-VNTR_{plus}, montre que Euro American-121 regroupe les isolats de la famille de Uganda II et de Ghana (annexe 5), alors que Euro American-12 rassemble principalement des isolats de la famille Cameroon. L'arbre phylogénétique MST de la figure 20 montre que Euro American-121 et Euro American-12 partagent un proche ancêtre. La lignée Euro American-121 serait plus ancienne que Euro American-12. Il y a une bonne présence des souches Euro-American-121 en Iran et une prévalence de Euro-American-12 en Afrique centrale et de l'Ouest. Cette observation mène à la conclusion qu'il y aurait eu une vague migratoire entre l'Asie de l'Ouest et l'Afrique de l'Ouest.

Cet échantillon d'Iran a également montré une forte prévalence de Euro American-212 dans le Nord du pays. La reclassification des isolats de MIRU-VNTR_{plus} et de SITVIT a montré que cette lignée correspondait à la famille spoligotypique S (annexe 5). Sola et al., en 2005, ont prouvé que la famille S avait une spécificité phylogéographique pour la Sicile et la Sardaigne. Cependant la forte prévalence de cette famille au Nord de l'Iran laisse supposer que la lignée Euro American-212 proviendrait du Moyen-Orient et se serait répandue sur la côte méditerranéenne, par une migration au néolithique, d'un groupe de personnes faisant partie de l'haplogroupe R1b.

En ce qui concerne la distribution des souches de MTC sur le continent américain, Hershberg et al., en 2008 l'explique par diverses migrations humaines récentes (datant de moins de 1500 ans), en provenance d'Europe de l'Ouest. Ainsi les sous-lignées Euro American se seraient dispersées sur le nouveau continent. Cependant, si l'on prête attention à la distribution de EAAI en Amérique, on constate que East Asian-17 prédomine en Amérique du Nord parmi les souches appartenant à EAAI avec 43,13% des souches EAAI. En Europe du Nord et de l'Ouest, c'est East African Indian-68 qui prédomine parmi les souches EAAI avec respectivement 37,71% et

31,51%. La forte prévalence de East Asian-17 parmi les isolats EAAI de l'Amérique du Nord rappelle celle de l'Asie de l'Est. Ceci est en faveur d'une migration des peuples de l'Est de l'Asie en Amérique du Nord, trajet qui est envisageable par le détroit de Béring.

Hershberg et al., en 2008 ont su superposer la distribution géographique des lignées MTC aux grands mouvements migratoires humains. Selon lui, l'histoire migratoire humaine qui ressort de la distribution des lignées de MTC est en adéquation avec l'hypothèse du out of Africa. Son étude fut réalisée à l'aide du séquençage de 89 gènes chez 108 souches distinctes. La distribution des lignées/sous-lignées MIRU-VNTR corrobore également cette hypothèse migratoire humaine du Out of Africa.

4.4. La capacité des loci VNTR à décrire les lignées phylogénétiques de MTC, sous-entendrait-elle un rôle de ces minisatellites chez les mycobactéries ?

4.4.1. Analyse bioinformatique des séquences de 12 loci MIRU-VNTR

Les minisatellites VNTRs sont des séquences réparties sur tout le génome du complexe *M. tuberculosis*, souvent localisées dans des régions intergéniques. Ces unités répétitives varient entre 40 et 100 pb. Les 12 loci MIRU-VNTRs décrit par Supply en 2000 sont des séquences VNTRs similaires entre elles (Supply et al., 2000).

Au cours d'une étude bioinformatique, plusieurs alignements de ces séquences avec le programme clustalW, ont permis d'identifier 4 sous-séquences dont les consensus sont :

- ACATGACCACCGCCGCGACGATGCACA (28 pb) ou C1
- GCGAAGCGATGAGGAG (16 pb) ou C2
- GAGCGGCGC (9 pb) ou C3
- GTGGGGGTACCGCCCGCTTGCGGGGGA (27 pb) ou C4

Par exemple, l'unité répétée de MIRU4 se compose de la sous séquence C1 (orange) suivie de C2 (violet), puis de C4 (rouge) et de C3 (vert) comme le montre la figure 27.

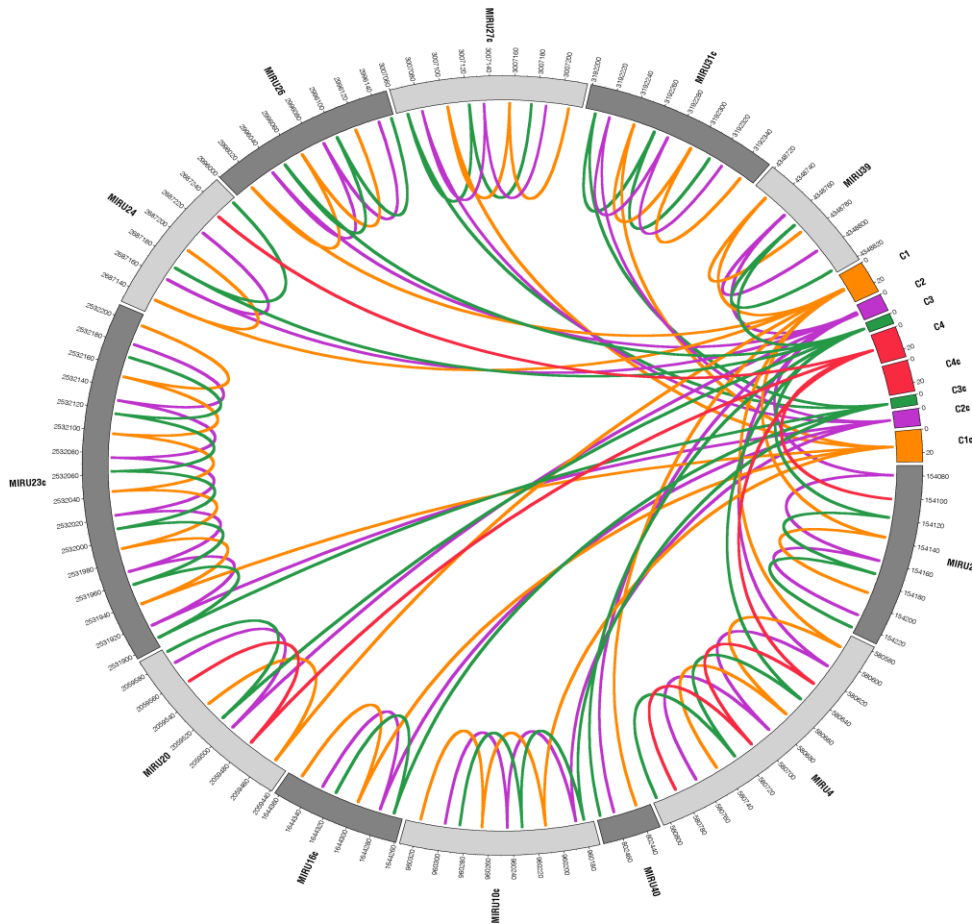


Figure 27. Composition de 12 loci VNTR (MIRU2, 4, 10, 16, 20, 23, 24, 26, 27, 31, 39, 40) chez H37Rv. Ces loci sont composés des sous-séquences C1, C2, C3, C4, ou de leurs complémentaires C1c, C2c, C3c, C4c. Image réalisée à l'aide du logiciel Circos (<http://circos.ca/>).

Comme le montre la figure 28, le MIRU4 chevauche les gènes *senX3* et *regX3*. Ce locus compte 3 répétitions chez *H37Rv*. Il présente plusieurs séquences Shine-Dalgarno, séquences propices à la fixation de la sous unité ribosomale 30S, sur l'ARNm pour le démarrage de la traduction. Les minisatellites MIRU-VNTR, lorsqu'ils sont en position intergénique, fournissent le codon « start » au gène en aval pour le démarrage de la traduction, et le codon « stop » au gène en amont (figure 28).

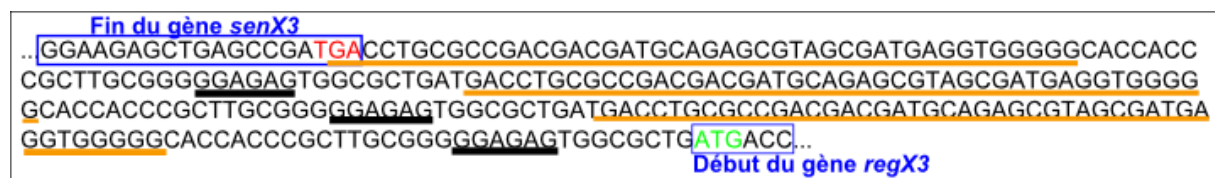


Figure 28. Séquence du MIRU4 dans H37Rv. Les portions de gènes sont encadrées en bleu. Le codon stop est indiqué en rouge, et le codon start en vert. Les séquences Shine-Dalgarno sont soulignées en noir.

4.4.2. Les loci MIRU-VNTRs seraient-ils les supports d'un mécanisme de régulation traductionnelle par le recodage du codon stop et le décalage de cadre de lecture chez les souches de MTC ?

4.4.2.1. Quel serait l'intérêt des codons stop portés par les minisatellites MIRU-VNTR ?

La particularité des VNTR à fournir des codons stop et la présence de ces minisatellites à proximité de gènes impliqués dans des activités redox sont des caractéristiques observées chez les éléments SECIS (**s**elenocysteine **i**nsertion **s**equences). A l'aide de l'outil *bSECISearch* (<http://genomics.unl.edu/bSECISearch/>) (Zhang and Gladyshev, 2005) pour détecter les éléments SECIS bactériens, on a pu mettre en évidence des éléments SECIS portés par les VNTRs. Par exemple la sous séquence 5'-UGACCUGCGCCGACGACGAUGCAGAGCGUAGCGAUGAGGUGGGGG-3' de MIRU4 prendrait cette conformation 2D in vivo (figure 29). La structure 2D de la figure 29 a été prédite par le logiciel *mfold* (<http://mfold.ma.albany.edu/?q=mfold>).

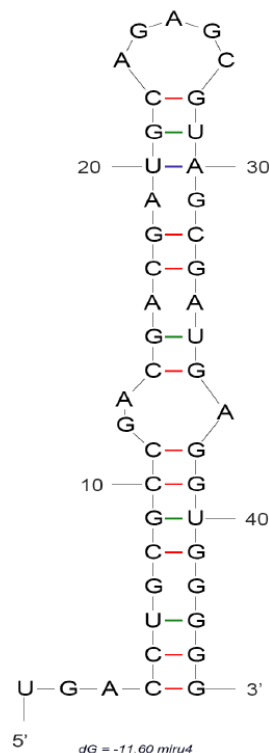


Figure 29. Élément SECIS du MIRU4 (prédiction faite avec le logiciel *bSECISearch*)

Cette structure de 45 pb très stable ($dG = -11,6$ kcal/mol) répond aux critères établis par (Zhang et Gladyshev, 2005), (figure 30).

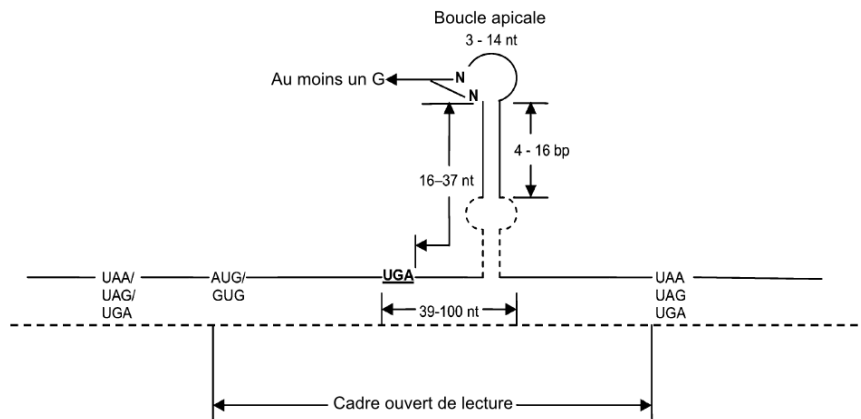


Figure 30. Élément SECIS consensus chez les bactéries (Zhang et Gladyshev, 2005)

4.4.2.2. Mécanisme d'insertion de la sélénocystéine (Sec) au cours de la traduction

Plusieurs modes de lecture du code génétique sont programmées par des signaux présents au niveau de l'ARNm. La présence d'élément secondaire en épingle à cheveux, par exemple, permet le recodage du codon stop « UGA » en sélénocystéine. Le sélénium est un micronutriment utilisé par beaucoup d'organismes. La forme biologique la plus courante du sélénium est la sélénocystéine, synthétisée à partir de la sérine portée par un ARN de transfert spécial (Figure 30). L'incorporation d'un sélénocystéine dans un gène requiert la présence de la sélénocystéine synthase (*selA*), de la sélénophosphate synthétase (*selD*), du tRNA^{Sec} (*selC*), d'un facteur d'élongation traductionnel spécialisé (*selB*) et d'un élément SECIS localisé immédiatement après l'UGA (figure 31).

selB joue un rôle important puisque qu'il se fixe à la fois au tRNA^{Sec} et à la structure SECIS, ce qui empêche la fixation du facteur de recyclage des ribosomes, chargé de libérer le peptide (figure 31).

Des analyses bioinformatiques réalisées dans cette étude, montrent que le génome mycobactérien possède toute la machinerie nécessaire à l'incorporation des sélénocystéines. En effet, à l'aide du serveur NCBI, il a été possible de vérifier la présence des gènes *selA*, *selB*, *selC* et *selD* chez les mycobactéries. Ces gènes sont quasiment toujours présents chez les mycobactéries atypiques (ou mycobactéries à croissance rapide) séquencées. Par exemple, chez *M. avium* 104 référence séquence : NC_008595.1), les éléments *selA*, *selB*, *selC* et *selD* sont portés respectivement par les loci MAV_0418, MAV_0417, MAV_0419 et MAV_0420. Grâce à un « tblastn » des protéines *selA*, *selB* et *selD*, on constate que ces protéines sont absentes chez les souches *M. tuberculosis* disponibles sur le serveur

NCBI. De même, un « blastn » de la séquence nucléique de *se/C* n'a donné aucun résultat significatif chez *M. tuberculosis*.

4.4.2.3. Structure et fonction de l'acide aminé sélénocystéine (Sec)

Les protéines ayant un résidu Sec ont été répertoriées dans les trois grands groupes d'êtres vivants : procaryotes, eucaryotes, et archaebactéries. Beaucoup de sélénoprotéines caractérisées sont des enzymes qui catalysent les réactions d'oxydo-réduction et qui contiennent Sec (sélénocystéine) dans leur site actif. Par exemple, la thioreduxine réductase qui est une flavoprotéine retrouvée chez les bactéries, les champignons ainsi que chez les plantes, présente un site rédox actif (disulfide/dithiol). Cette protéine présente un motif conservé contenant Sec qui forme un site redox actif (Arner and Holmgren, 2000). De plus, on peut remarquer la présence du VNTR Mtub39 à proximité d'une disulfide réductase.

Structurellement, l'élément Sec est identique à la cystéine (Cys), sauf que le soufre fut remplacé par le sélénium. Le sélénium et le soufre font parties de la même famille des chalcogènes (caractérisés par une forte tendance à capter des électrons ou à former des liaisons covalentes) selon le tableau périodique. Ces deux éléments partagent de nombreuses propriétés chimiques. Sec présente un avantage fonctionnel par rapport à Cys car le groupement selenol (R-SeH) est davantage ionisé que le groupement thiol (R-SH) de Cys à pH physiologique. Lorsque Sec est remplacé par Cys, l'activité catalytique d'une sélénoenzyme est drastiquement réduite.

4.4.2.4. Compétition entre l'insertion de Sec et la terminaison

L'efficacité de l'insertion de Sec dans *E. coli* est assez faible par rapport à l'incorporation d'autres acides aminés. Cette inefficacité n'est pas uniquement due à la limitation des composants de la machinerie d'intégration de Sec, vu que l'augmentation de la concentration des acteurs clés (Selb, tRNA^{Sec} ou Sec synthétase) augmente l'intégration de Sec seulement d'un facteur deux à trois.

La synthèse des sélénoprotéines est stimulée lorsque la concentration en tRNA^{Sec} est augmentée, ce qui suggère que cette molécule est un facteur limitant. La sérine est remplacée par Sec uniquement lorsque le sélénium est disponible pour synthétiser les donneurs sélénophosphates. Si le sélénium est limité, il n'y a pas de

tRNA^{Sec} disponible pour le ribosome, ce qui entraîne le décodage d'UGA en un signal de terminaison de la traduction. Donc, l'insertion de Sec et la terminaison traductionnelle sont des événements compétitifs.

Chez les procaryotes, la terminaison est catalysée par deux facteurs de libération (RFs, pour release factors) qui reconnaissent le codon stop. RF1 reconnaît UAG et UAA, alors que RF2 reconnaît UGA et UAA. RF2 et tRNA^{Sec} sont donc en compétition pour le même site de liaison sur le ribosome. Le changement de la concentration cellulaire de ces deux facteurs module le décodage d'UGA comme Sec ou Stop.

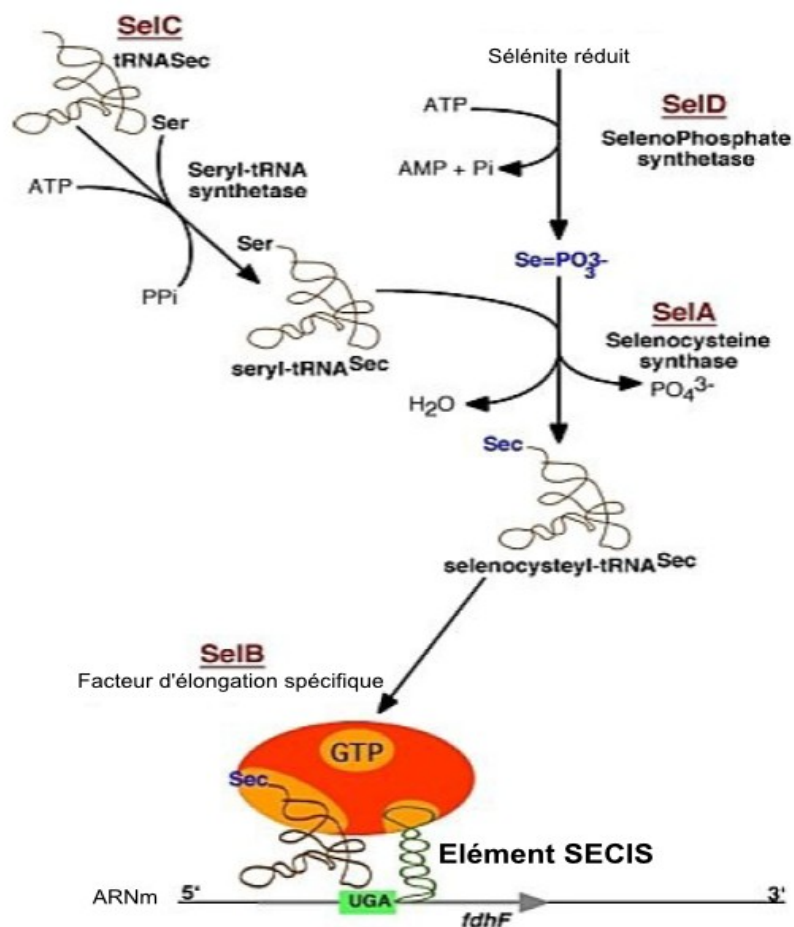


Figure 31. Mécanisme d'incorporation de l'élément Sec. Image du site : <http://www.uni-frankfurt.de/fb/fb15/institute/inst-3-mol-biowiss/AK-Rother/research.html>

4.4.2.5. Quel serait l'intérêt des séquences Shine-Dalgarno portées par les VNTRs ? (réponse apportée par l'article Namy et al., 2004)

La séquence Shine-Dalgarno (SD) ainsi que les structures « tige-boucle » (ou pseudonot) sont impliquées dans des phénomènes de changement de cadre de

lecture au cours de la traduction protéique. Le ribosome peut glisser d'une base en amont (frameshift -1) ou en aval (frameshift +1). Cet événement, n'arrivant pas de façon systématique, permet de réguler la concentration de certaines protéines. Par exemple, le gène *dnaX* de *E. coli* montre un bel exemple de ce mécanisme. Ce gène code à la fois pour les sous-unités γ et τ de la polymérase III. La synthèse de la sous-unité γ est due à un -1 frameshift conduisant à un codon stop prématuré, alors que τ , ayant une séquence plus longue, est traduite sur l'ensemble de l'ORF (Open Reading Frame ou cadre ouvert de lecture) sans changement de cadre de lecture. Le glissement se produit au niveau de la séquence 5'-A AAA AAG-3' (figure 32).

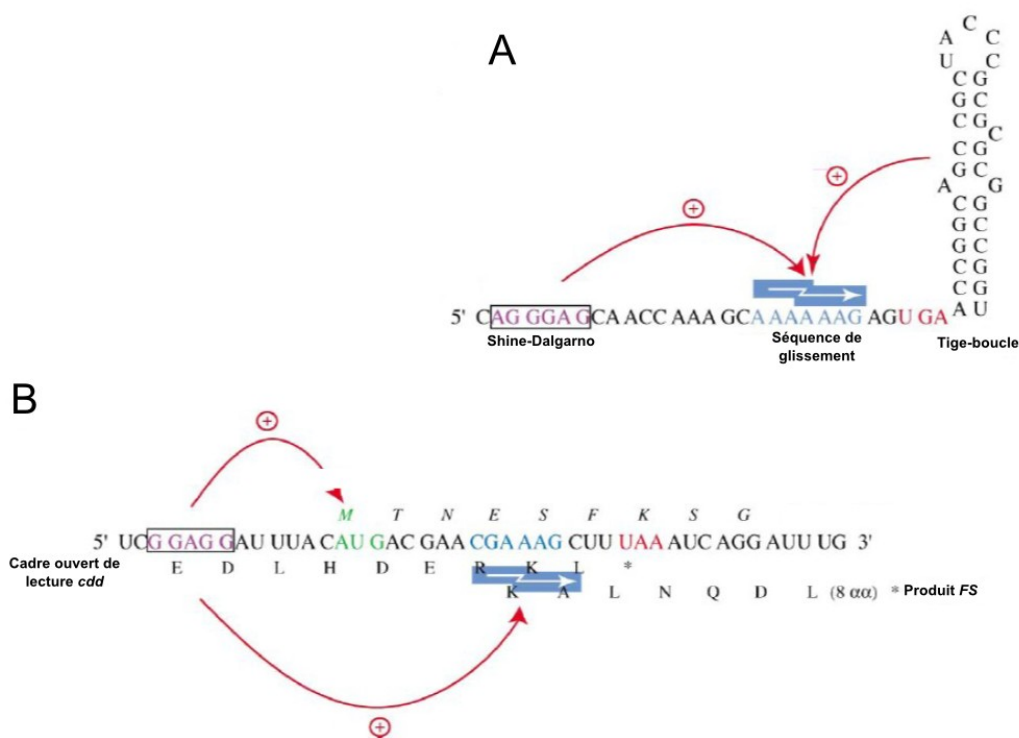


Figure 32. Exemple de frameshift -1 chez les procaryotes. A) frameshift dans le gène *dnaX* de *E. coli*. B) frameshift dans le gène *cdd* de *B. subtilis* (Namy et al., 2004).

Celui-ci entraîne un changement de lecture de 5'-A AAA AAG-3' en 5'-AAA AAA G-3' (figure 32B). La séquence SD, la séquence de glissement ainsi que la structure en épingle à cheveux, induisent une tension dans l'ARNm qui provoque un repositionnement du ribosome. De même chez *B. subtilis* avec la cytidine desaminase (*cdd*), on observe un changement de cadre de lecture à cause de la présence d'une SD et d'un codon start à proximité d'un codon stop (figure 32C). Chez *E. coli* on observe un frameshift +1 sur le gène *prfB* codant pour une RF2

permettant une autorégulation du RF2 (Namy et al., 2004). Cette régulation affecte donc l'incorporation de Sec. Il est important de souligner la présence de VNTRs à proximité des séquences codant pour une RF (VNTR : ETR-E), pour *dnaQ* qui code pour une sous unité de la polymérase III (VNTR : QUB 4156c), et pour une DNA Polymérase I (VNTR : Mtub29).

4.4.2.6. Conclusion sur une éventuelle fonction des minisatellites MIRU-VNTR chez les souches MTC

L'implication des MIRU-VNTR dans un mécanisme de régulation traductionnelle par le recodage du codon Stop et le décalage de cadre de lecture chez les souches de MTC reste hypothétique. L'analyse bioinformatique réalisée sur ces séquences et le résultat des recherches bibliographiques effectuées sur le sujet permettent seulement d'émettre l'hypothèse. On peut seulement affirmer que les VNTRs peuvent prendre une conformation 2D obéissant aux caractéristiques des éléments SECIS qui permettent l'insertion de Sec dans les protéines et que ces minisatellites portent également des séquences Shine-Dalgarno qui pourraient être impliquées dans le repositionnement des ribosomes au cours de la traduction.

Pour vérifier si ces événements ont réellement lieu chez les bactéries du complexe *Mycobacterium tuberculosis*, il est nécessaire de procéder à des études supplémentaires.

RÉFÉRENCES

- Abadia E, Zhang J, dos Vultos T, Ritacco V, Kremer K, et al. (2010) Resolving lineage assignment on *Mycobacterium tuberculosis* clinical isolates classified by spoligotyping with a new high-throughput 3R SNPs based method. *Infect Genet Evol* 10: 1066–1074.
- Agerton TB, Valway SE, Blinkhorn RJ, Shilkret KL, Reves R, et al. (1999) Spread of strain W, a highly drug-resistant strain of *Mycobacterium tuberculosis*, across the United States. *Clin Infect Dis* 29: 85-92; discussion 93-85.
- Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, Gey van Pittius NC. Novel *Mycobacterium tuberculosis* complex pathogen, *M. mungi*. *Emerg Infect Dis*. 2010 Aug;16(8):1296-9. Erratum in: *Emerg Infect Dis*. 2010 Dec;16(12):2024.
- Al-Hajoj SA, Zozio T, Al-Rabiah F, Mohammad V, Al-Nasser M, et al. (2007) First insight into the population structure of *Mycobacterium tuberculosis* in Saudi Arabia. *J Clin Microbiol* 45: 2467–2473.
- Almeida D, Rodrigues C, Ashavaid TF, Lalvani A, Udwardia ZF, 359 et al. (2005) High incidence of the Beijing genotype among multidrug-resistant isolates of *Mycobacterium tuberculosis* in a tertiary care center in Mumbai, India. *Clin Infect Dis* 40: 881-886.
- Anilkumar AK, Madhavalatha GK, Paul LK, Radhakrishnan I, Kumar RA, Mundayoor S. Standardization and evaluation of a tetraplex polymerase chain reaction to detect and differentiate *Mycobacterium tuberculosis* complex and nontuberculous *Mycobacteria*-a retrospective study on pulmonary TB patients. *Diagn Microbiol Infect Dis*. 2012 Mar;72(3):239-47. Epub 2011 Dec 29.
- Aranaz A, Cousins D, Mateos A, Domínguez L. Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz et al. 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. *Int J Syst Evol Microbiol*. 2003 Nov;53(Pt 6):1785-9.
- Aranaz A, Liébana E, Gómez-Mampaso E, Galán JC, Cousins D, Ortega A, Blázquez J, Baquero F, Mateos A, Suárez G, Domínguez L. *Mycobacterium tuberculosis* subsp. *caprae* subsp. nov.: a taxonomic study of a new member of the *Mycobacterium tuberculosis* complex isolated from goats in Spain. *Int J Syst Bacteriol*. 1999 Jul;49 Pt 3:1263-73.
- Aranaz A., Liebana E., Mateos A., Dominguez L., Cousins S. ~ Restriction Fragment Length Polymorphism and spacer oligonucleotide typing: a comparative analysis of fingerprinting strategies for *Mycobacterium bovis*. *Vet. Microbiol.*, 1998,61:311-324.
- Aranaz A., Liebana E., Mateos A., Dominguez L., Vidal D., Domingo M., Gonzolez O., Rodriguez- Ferri E.F., Bunschoten A.E., van Embden J.D.A. and Cousins D. ~ Spacer oligonucleotide typing of *Mycobacterium bovis* strains from cattle and other animals: a tool for studying epidemiology of tuberculosis. *J. Clin. Microbiol.*, 1996, 34:2734-2740,
- Arnér ES, Holmgren A. Physiological functions of thioredoxin and thioredoxin reductase. *Eur J Biochem*. 2000 Oct;267(20):6102-9.
- Asgharzadeh, M, HS Kafil, AA Roudsary, GR Hanifi. 2011. Tuberculosis transmission in Northwest of Iran: using MIRU-VNTR, ETR-VNTR and IS6110-RFLP methods. *Infect* 11:124-131. Epub 2010 Oct 2014.
- Asselineau J, Lederer E. [The chemical constitution of the mycolic acids of two virulent human strains of *Mycobacterium tuberculosis*]. *Biochim Biophys Acta*. 1951 May;7(1):126-45. Undetermined Language.
- Baldeviano-Vidalón, G.C., Quispe-Torres, N., Bonilla-Asalde, C., Gastiaburú-Rodríguez, D., Pro-Cuba, J.E., Llanos-Zavalaga, F., 2005. Multiple infection with resistant and sensitive *M. Tuberculosis* strains during treatment of pulmonary tuberculosis patients. *Int. J. Tuberc. Lung Dis*. 9, 1155–1160,
- Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, Wilson T, Collins D, de Lisle G, Jacobs WR Jr. inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science*. 1994 Jan 14;263(5144):227-30,
- Beggs ML, Eisenach KD, Cave MD. Mapping of IS6110 insertion sites in two epidemic strains of *Mycobacterium tuberculosis*. *J Clin Microbiol*. 2000 Aug;38(8):2923-8.
- Bhanu NV, van Soolingen D, van Embden JD, Dar L, Pandey RM, Seth P. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. *Tuberculosis (Edinb)*. 2002;82(2-3):105-12.
- Bickhart DM, Gogarten JP, Lapierre P, Tisa LS, Normand P, Benson DR. Insertion sequence content

- reflects genome plasticity in strains of the root nodule actinobacterium *Frankia*. *BMC Genomics*. 2009 Oct 12;10:468.
- Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. *Trends Microbiol*. 2002 Jan;10(1):45-52.
- Bifani PJ, Plikaytis BB, Kapur V, Stockbauer K, Pan X, Lutfey ML, Moghazeh SL, Eisner W, Daniel TM, Kaplan MH, Crawford JT, Musser JM, Kreiswirth BN. Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. *JAMA*. 1996 Feb 14;275(6):452-7.
- Billamas P, Smittipat N, Juthayothin T, Thong-On A, Yamada N, et al. (2007) Evolution of some variable-number tandem repeat loci among a group of Beijing strains of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 87: 498–501.
- Brinig MM, Cummings CA, Sanden GN, Stefanelli P, Lawrence A, Relman DA. Significant gene order and expression differences in *Bordetella pertussis* despite limited gene content variation. *J Bacteriol*. 2006 Apr;188(7):2375-82.
- Boddinghaus B, Rogall T, Flohr T, Blocker H, Bottger EC (1990) Detection and identification of mycobacteria by amplification of rRNA. *J Clin Microbiol* 28: 1751-1759.
- Brodin, P, K Eiglmeier, M Marmiesse, A Billault, T Garnier, S Niemann, ST Cole, R Brosch. 2002. Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. *Infect Immun*. 70:5568-5578.
- Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A*. 2002 Mar 19;99(6):3684-9.
- Brosch R, Philipp WJ, Stavropoulos E, Colston MJ, Cole ST, Gordon SV. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. *Infect Immun*. 1999 Nov;67(11):5768-74. Erratum in: *Infect Immun* 2000 Jan;68(1):427.
- Brozna, J. P., M. Horan, J. M. Rademacher, K. M. Pabst, and M. J. Pabst. 1991. Monocyte responses to sulfolipid from *Mycobacterium tuberculosis*: inhibition of priming for enhanced release of superoxide, associated with increased secretion of interleukin-1 and tumor necrosis factor alpha and altered protein phosphorylation. *Infect. Immun*. 59:2542-2548.
- Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajj SA, Allix C, Aristimuño L, Arora J, Baumanis V, Binder L, Cafrune P, Cataldi A, Cheong S, Diel R, Ellermeier C, Evans JT, Fauville-Dufaux M, Ferdinand S, Garcia de Viedma D, Garzelli C, Gazzola L, Gomes HM, Gutierrez MC, Hawkey PM, van Helden PD, Kadival GV, Kreiswirth BN, Kremer K, Kubin M, Kulkarni SP, Liens B, Lillebaek T, Ho ML, Martin C, Martin C, Mokrousov I, Narvskaja O, Ngeow YF, Naumann L, Niemann S, Parwati I, Rahim Z, Rasolofo-Razanamparany V, Rasolonavalona T, Rossetti ML, Rüscher-Gerdes S, Sajduda A, Samper S, Shemyakin IG, Singh UB, Somoskovi A, Skuere RA, van Soolingen D, Streicher EM, Suffys PN, Tortoli E, Tracevska T, Vincent V, Victor TC, Warren RM, Yap SF, Zaman K, Portaels F, Rastogi N, Sola C. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol*. 2006 Mar 6;6:23.
- Cadmus S, Palmer S, Okker M, Dale J, Gover K, Smith N, Jahans K, Hewinson RG, Gordon SV. Molecular analysis of human and bovine tubercle bacilli from a local setting in Nigeria. *J Clin Microbiol*. 2006 Jan;44(1):29-34.
- Cadmus, S. I. B., A. A. Atsanda, S. O. Oni, and E. E. U. Akang. 2004. Bovine tuberculosis in one cattle herd in Ibadan in Nigeria. *Vet. Med. Czech*. 49:406–412.
- Cadmus, S. I. B., H. K. Adesokan, A. F. Adepoju, and E. B. Otesile. 2008. Zoonotic risks and transmission of *Mycobacteria* species from cows' milk and slaughtered cattle to man in Ibadan: role of butchers. *Niger. Vet. J*. 291:30–39.
- Caimi, K, MI Romano, A Alito, M Zumarraga, F Bigi, A Cataldi. 2001. Sequence analysis of the direct repeat region in *Mycobacterium bovis*. *J Clin Microbiol* 39:1067-1072.
- Castets M, Boisvert H, Grumbach F, Brunel M, and Rist N. Tuberculosis bacilli of the African type: preliminary note. *Rev. Tuberc. Pneumol*. 1968, 32:179-184.
- Cavanagh R, Begon M, Bennett M, Ergon T, Graham IM, De Haas PE, Hart CA, Koedam M, Kremer K, Lambin X, Roholl P, Soolingen Dv D. *Mycobacterium microti* infection (vole tuberculosis) in wild rodent populations. *J Clin Microbiol*. 2002 Sep;40(9):3281-5.
- Comas I, Homolka S, Niemann S, Gagneux S (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current

- methodologies. PLoS One 4: e7815.
- Cousins DV, Bastida R, Cataldi A, Quse V, Redrobe S, Dow S, Duignan P, Murray A, Dupont C, Ahmed N, Collins DM, Butler WR, Dawson D, Rodríguez D, Loureiro J, Romano MI, Alito A, Zumarraga M, Bernardelli A. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int J Syst Evol Microbiol.* 2003 Sep;53(Pt 5):1305-14.
- Cruciani, F, B Trombetta, D Sellitto, A Massaia, G Destro-Bisol, E Watson, E Beraud Colomb, JM Dugoujon, P Moral, R Scozzari. 2010, Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *Eur J Hum Genet.* 18:800-807. Epub 2010 Jan 2016.
- Darwin, C. (1997, éd. or. 1859) *L'Origine des espèces*, Flammarion.
- Das S, Paramasivan CN, Lowrie DB, Prabhakar R, Narayanan PR (1995) IS6110 restriction fragment length polymorphism typing of clinical isolates of *Mycobacterium tuberculosis* from patients with pulmonary tuberculosis in Madras, south India. *Tuber Lung Dis* 76: 550-554.
- Das, S., Narayanan, S., Hari, L., Mohan, N.S., Somasundaram, S., Selvakumar, N., Narayanan, P.R., 2004. Simultaneous infection with multiple strains of *Mycobacterium tuberculosis* identified by restriction fragment length polymorphism analysis. *Int. J. Tuberc. Lung Dis.* 8, 267–270,
- de Jong BC, Antonio M, Awine T, Ogungbemi K, de Jong YP, Gagneux S, DeRiemer K, Zozio T, Rastogi N, Borgdorff M, Hill PC, Adegbola RA. Use of spoligotyping and large sequence polymorphisms to study the population structure of the *Mycobacterium tuberculosis* complex in a cohort study of consecutive smear-positive tuberculosis cases in The Gambia. *J Clin Microbiol.* 2009 Apr;47(4):994-1001. Epub 2009 Feb 4.
- de Jong BC, Antonio M, Gagneux S. *Mycobacterium africanum*--review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis.* 2010 Sep 28;4(9):e744. Review.
- Delespesse, G., C. E. Demeure, L. P. Yang, Y. Ohshima, D. G. Byun, and U. Shu. 1997. In vitro maturation of naive human CD41 T lymphocytes into Th1, Th2 effectors. *Int. Arch. Allergy Immunol.* 113(1–3):157–159.
- Demay, C., B. Liens, T. Burguière, V. Hill, D. Couvin, J. Millet, I. Mokrousov, C. Sola, T. Zozio, N. Rastogi. 2012. SITVITWEB – A publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol.* Feb 17. [Epub ahead of print]
- Devallois A., Horgen L., Sola C. and Rastogi N. Le typage moléculaire des mycobactéries. *Path.Biol.*, 1998, 46:625-636.
- Epardeau B, Levagueresse R, Fabre M. 1982. L'antibiogramme du bacille de la tuberculose en milieu liquide. Technique, premiers résultats, intérêt de la méthode. *Médecine et armées* 10-9.
- Eldholm V, Matee M, Mfinanga SG, Heun M, Dahle UR. A first insight into the genetic diversity of *Mycobacterium tuberculosis* in Dar es Salaam, Tanzania, assessed by spoligotyping. *BMC Microbiol.* 2006 Sep 13;6:76.
- Fang, Z, KJ Forbes. 1997. A *Mycobacterium tuberculosis* IS6110 preferential locus (ipl) for insertion into the genome. *J. Clin. Microbiol.* 35:479-481.
- Fang, Z, N Morrison, B Watt, C Doig, KJ Forbes. 1998. IS6110 transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. *J Bacteriol* 180:2102-2109.
- Ferdinand S, Sola C, Chanteau S, Ramarokoto H, Rasolonavalona T, et al. (2005) A study of spoligotyping-defined *Mycobacterium tuberculosis* clades in relation to the origin of peopling and the demographic history in Madagascar. *Infect Genet Evol* 5: 340–348.
- Filliol I, Driscoll JR, Van Soolingen D, Kreiswirth BN, Kremer K, Valétudie G, Anh DD, Barlow R, Banerjee D, Bifani PJ, Brudey K, Cataldi A, Cooksey RC, Cousins DV, Dale JW, Dellagostin OA, Drobniewski F, Engelmann G, Ferdinand S, Gascoyne-Binzi D, Gordon M, Gutierrez MC, Haas WH, Heersma H, Källénus G, Kassa-Kelembho E, Koivula T, Ly HM, Makristathis A, Mamina C, Martin G, Moström P, Mokrousov I, Narbonne V, Narvskaya O, Nastasi A, Niobe-Eyangoh SN, Pape JW, Rasolofo-Razanamparany V, Ridell M, Rossetti ML, Stauffer F, Suffys PN, Takiff H, Texier-Maugein J, Vincent V, De Waard JH, Sola C, Rastogi N. Global distribution of *Mycobacterium tuberculosis* spoligotypes. *Emerg Infect Dis.* 2002 Nov;8(11):1347-9.
- Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbón MH, Bobadilla del Valle M, Fyfe J, García-García L, Rastogi N, Sola C, Zozio T, Guerrero MI, León CI, Crabtree J, Angiuoli S, Eisenach KD, Durmaz R, Joloba ML, Rendón A, Sifuentes-Osornio J, Ponce de León A, Cave MD, Fleischmann R, Whittam TS, Alland D. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal

- standard SNP set. *J Bacteriol.* 2006 Jan;188(2):759-72. Erratum in: *J Bacteriol.* 006 Apr;188(8):3162-3.
- Finken M, Kirschner P, Meier A, Wrede A, Böttger EC. Molecular basis of streptomycin resistance in *Mycobacterium tuberculosis*: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. *Mol Microbiol.* 1993 Sep;9(6):1239-46.
- Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E, Kolonay JF, Nelson WC, Umayam LA, Ermolaeva M, Salzberg SL, Delcher A, Utterback T, Weidman J, Khouri H, Gill J, Mikula A, Bishai W, Jacobs Jr WR Jr, Venter JC, Fraser CM. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol.* 2002 Oct;184(19):5479-90, PubMed PMID: 12218036; PubMed Central PMCID: PMC135346.
- Frothingham R, Meeker-O'Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology.* 1998 May;144 (Pt 5):1189-96.
- Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, Nicol M, Niemann S, Kremer K, Gutierrez MC, Hilty M, Hopewell PC, Small PM. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 2006 Feb 21;103(8):2869-73. Epub 2006 Feb 13.
- Gagneux S, Small PM (2007) Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 7: 328–337.
- Ganguly N, Siddiqui I, Sharma P. Role of *M. tuberculosis* RD-1 region encoded secretory proteins in protective response and virulence. *Tuberculosis (Edinb).* 2008 Nov;88(6):510-7. Epub 2008 Jul 21.
- Garg A, Ayyagari A. Reference to the mutations in the embB locus among Korean clinical isolates of *Mycobacterium tuberculosis* resistant to ethambutol. *Yonsei Med J.* 2002 Aug;43(4):556; author reply 556.
- Gibson AL, Huard RC, Gey van Pittius NC, Lazzarini LC, Driscoll J, et al. (2008) Application of sensitive and specific molecular methods to uncover global dissemination of the major RDRio Sublineage of the Latin American-Mediterranean *Mycobacterium tuberculosis* spoligotype family. *J Clin Microbiol* 46: 1259–1267.
- Glynn JR, Alghamdi S, Mallard K, McNerney R, Ndlovu R, Munthali L, Houben RM, Fine PE, French N, Crampin AC. Changes in *Mycobacterium tuberculosis* genotype families over 20 years in a population-based study in Northern Malawi. *PLoS One.* 2010 Aug 17;5(8):e12259.
- Gordon SV, Heym B, Parkhill J, Barrell B, Cole ST. New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. *Microbiology.* 1999 Apr;145 (Pt 4):881-92.
- Goren, M. B., P. D'Arcy Hart, M. R. Young, and J. A. Armstrong. 1976. Prevention of phagosome-lysosome fusion in cultured macrophages by sulfatides of *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* 73:2510-2514.
- Goyal M., Lawn S., Afful B., Acheampong J.W., Griffin G. and Shaw R. ~ Spoligotyping in molecular epidemiology of tuberculosis in Ghana. *J. Infect.*, 1999, 38:171-175.
- Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, Supply P, Vincent V. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog.* 2005 Sep;1(1):e5. Epub 2005 Aug 19.
- Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, Katoch VM, Vincent V, Locht C, Supply P. Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerg Infect Dis.* 2006 Sep;12(9):1367-74.
- Hanekom M, van der Spuy GD, Gey van Pittius NC, McEvoy CR, Hoek KG, Ndabambi SL, Jordaan AM, Victor TC, van Helden PD, Warren RM. Discordance between mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing and IS6110 restriction fragment length polymorphism genotyping for analysis of *Mycobacterium tuberculosis* Beijing strains in a setting of high incidence of tuberculosis. *J Clin Microbiol.* 2008 Oct;46(10):3338-45.
- Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. 1970 *Biometrika* 57 : 97-109.
- Helal ZH, Ashour MS, Eissa SA, Abd-Elatef G, Zozio T, et al. (2009) Unexpectedly high proportion of ancestral Manu genotype *Mycobacterium tuberculosis* strains cultured from tuberculosis patients in Egypt. *J Clin Microbiol* 47: 2794–2801.
- Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, Roach JC, Kremer K, Petrov DA, Feldman MW, Gagneux S. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol.* 2008 Dec 16;6(12):e311.

- Ho, TB, BD Robertson, GM Taylor, RJ Shaw, DB Young. 2000, Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. *Yeast*. 17:272-282.
- Honore N. and S. Cole. Streptomycin resistance in mycobacteria *Antimicrob*. 1993. Agents Chemother. 38:238-242.
- Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science*. 2010 Jan 8;327(5962):167-70, Review.
- Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001 Aug;17(8):754-5.
- Jacobs RF. Multiple-drug-resistant tuberculosis. *Clin Infect Dis*. 1994 Jul;19(1):1-8; quiz 9-10, Review.
- Jad WATFA, Frédéric MICHEL. Tuberculose uro-génitale. *Progrès en Urologie* (2005), 15, 602-603
- Jansen, R, JD Embden, W Gaastra, LM Schouls. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. 43:1565-1575.
- Joseph BV, Soman S, Radhakrishnan I, Madhavilatha GK, Paul LK, et al. (2009) Drug resistance in *Mycobacterium tuberculosis* isolates from tuberculosis patients in Kerala, India. *Int J Tuberc Lung Dis* 13: 494-499.
- Joseph MR, Shoby CT, Amma GR, Chauhan LS, Paramasivan CN (2007) Surveillance of anti-tuberculosis drug resistance in Ernakulam District, Kerala State, South India. *Int J Tuberc Lung Dis* 11: 443-449.
- Kallenius G, Koivula T, Ghebremichael S, Hoffner SE, Norberg R, et al. (1999) Evolution and clonal traits of *Mycobacterium tuberculosis* complex in Guinea-Bissau. *J Clin Microbiol* 37: 3872–3878.
- Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol*. 1997 Apr;35(4):907-14
- Kapur V, Whittam TS, Musser JM. Is *Mycobacterium tuberculosis* 15,000 years old? *J Infect Dis*. 1994 Nov;170(5):1348-9. PubMed PMID: 7963745.
- Karlson AG, Lessel EF. "*Mycobacterium bovis* nom. nov.". *International Journal of Systematic Bacteriology*. 1970, 20 (3): 273–282.
- Kayser, M. 2010, The human genetic history of Oceania: near and remote views of dispersal. *Curr* 20:R194-201.
- Kiers A, Klarenbeek A, Mendelits B, Van Soolingen D, Koëter G. Transmission of *Mycobacterium pinnipedii* to humans in a zoo with marine mammals. *Int J Tuberc Lung Dis*. 2008 Dec;12(12):1469-73.
- Klevytska, AM, LB Price, JM Schupp, PL Worsham, J Wong, P Keim. 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol*. 39:3179-3185.
- Koeck, JL, BM Njanpop-Lafourcade, S Cade, E Varon, L Sangare, S Valjevac, G Vergnaud, C Pourcel. 2005. Evaluation and selection of tandem repeat loci for *Streptococcus pneumoniae* MLVA strain typing. *BMC Microbiol*. 5:66.
- Koivula T, Ekman M, Leitner T, Lofdahl S, Ghebremicahel S, et al. (2004) Genetic characterization of the Guinea-Bissau family of *Mycobacterium tuberculosis* complex strains. *Microbes Infect* 6: 272–278.
- Kirillov MYu, Shumakov YuL, Nechaeva EV, Butcher S, Sinjashina L, Runeberg K, Romantschuk M, Karataev GI. Repeated sequences isolated from *Bordetella pertussis* induce DNA rearrangements and deletions at high frequency. *Gene*. 1995 Dec 1;166(1):111-6.
- Lari, N, L Rindi, C Garzelli. 2001. Identification of one insertion site of IS6110 in *Mycobacterium tuberculosis* H37Ra and analysis of the RvD2 deletion in M. tuberculosis clinical isolates. *J Med Microbiol*. 50:805-811.
- Lazzarini LC, Huard RC, Boechat NL, Gomes HM, Oelemann MC, Kurepina N, Shashkina E, Mello FC, Gibson AL, Virginio MJ, Marsico AG, Butler WR, Kreiswirth BN, Suffys PN, Lapa E Silva JR, Ho JL. Discovery of a novel *Mycobacterium tuberculosis* lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. *J Clin Microbiol*. 2007 Dec;45(12):3891-902. Epub 2007 Sep 26.
- Lazzarini LC, Rosenfeld J, Huard RC, Hill V, Lapa E Silva JR, Desalle R, Rastogi N, Ho JL. *Mycobacterium tuberculosis* spoligotypes that may derive from mixed strain infections are revealed by a novel computational approach. *Infect Genet Evol*. 2011 Sep 5.
- Legrand, E, I Filliol, C Sola, N Rastogi. 2001. Use of Spoligotyping To Study the Evolution of the Direct Repeat Locus by IS6110 Transposition in *Mycobacterium tuberculosis*. *J Clin Microbiol* 39:1595-1599.

- Lewis PO. Phylogenetic systematics turns over a new leaf. *Trends Ecol Evol.* 2001 Jan 1;16(1):30-37.
- Liao CH, Lai CC, Ding LW, Hou SM, Chiu HC, Chang SC, Hsueh PR. Skin and soft tissue infection caused by non-tuberculous mycobacteria. *Int J Tuberc Lung Dis.* 2007 Jan;11(1):96-102.
- Majumder, PP. 2010, The human genetic history of South Asia. *Curr* 20:R184-187.
- Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol.* 1996 Mar;178(5):1274-82.
- McHugh TD, Batt SL, Shorten RJ, Gosling RD, Uiso L, Gillespie SH. *Mycobacterium tuberculosis* lineage: a naming of the parts. *Tuberculosis(Edinb).* 2005 May;85(3):127-36.
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of state calculations by fast computing machines. 1953 *J. Chemical Physics*, 21 : 1087-1091.
- Metropolis N, Ulam S. The Monte Carlo method. *J Am Stat Assoc.* 1949 Sep;44(247):335-41.
- Michel Tibayrenc. *Encyclopedia of infectious diseases: modern methodologies.* Hoboken (N.J.) : Wiley-Liss, cop. 2007.
- Milan SJ, Hauge KA, Kurepina NE, Lofy KH, Goldberg SV, Narita M, Nolan CM, McElroy PD, Kreiswirth BN, Cangelosi GA. Expanded geographical distribution of the N family of *Mycobacterium tuberculosis* strains within the United States. *J Clin Microbiol.* 2004 Mar;42(3):1064-8.
- Miltgen J, Morillon M, Koeck JL, Varnerot A, Briant JF, Nguyen G, Verrot D, Bonnet D, Vincent V. Two cases of pulmonary tuberculosis caused by *Mycobacterium tuberculosis* subsp *canetti*. *Emerg Infect Dis.* 2002 Nov;8(11):1350-2.
- Mistry NF, Iyer AM, D'souza DT, Taylor GM, Young DB, Antia NH. Spoligotyping of *Mycobacterium tuberculosis* isolates from multiple-drug-resistant tuberculosis patients from Bombay, India. *J Clin Microbiol.* 2002 Jul;40(7):2677-80,
- Mokrousov I, Valcheva V, Sovhozova N, Aldashev A, Rastogi N, Isakova J. 2009. Penitentiary population of *Mycobacterium tuberculosis* in Kyrgyzstan: exceptionally high prevalence of the Beijing genotype and its Russia-specific subtype. *Infect Genet Evol.* 9:1400-5.
- Mokrousov, I, HM Ly, T Otten, NN Lan, B Vyshnevskiy, S Hoffner, O Narvskaya. 2005. Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. *Genome Res.* 15:1357-1364. Epub 2005 Sep 1316.
- Mostowy S, Onipede A, Gagneux S, Niemann S, Kremer K, Desmond EP, Kato-Maeda M, Behr M. Genomic analysis distinguishes *Mycobacterium africanum*. *J Clin Microbiol.* 2004 Aug;42(8):3594-9.
- Myres, NM, S Rootsi, AA Lin, et al. 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur* 19:95-101. Epub 2010 Aug 2025.
- Namouchi A, Karboul A, Mhenni B, Khabouchi N, Haltiti R, et al. (2008) Genetic profiling of *Mycobacterium tuberculosis* in Tunisia: predominance and evidence for the establishment of a few genotypes. *J Med Microbiol* 57: 864–872.
- Namy O, Rousset JP, Naphtine S, Brierley I. Reprogrammed genetic decoding in cellular gene expression. *Mol Cell.* 2004 Jan 30;13(2):157-68.
- Narayanan S, Gagneux S, Hari L, Tsolaki AG, Rajasekhar S, Narayanan PR, Small PM, Holmes S, Deriemer K. Genomic interrogation of ancestral *Mycobacterium tuberculosis* from south India. *Infect Genet Evol.* 2008 Jul;8(4):474-83. Epub 2007 Sep 29. PubMed PMID: 18024233.
- Niobe-Eyangoh SN, Kuaban C, Sorlin P, Cunin P, Thonnon J, et al. (2003) Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. *J Clin Microbiol* 41: 2547–2553
- Niobe-Eyangoh SN, Kuaban C, Sorlin P, Thonnon J, Vincent V, Gutierrez MC. Molecular characteristics of strains of the cameroon family, the major group of *Mycobacterium tuberculosis* in a country with a high prevalence of tuberculosis. *J Clin Microbiol.* 2004 Nov;42(11):5029-35.
- Ohnishi T, Kusumoto S, Yamaguchi S, Ohki Y, Satou M, Sugiyama T, Shirai T, Nakashima M, Yamaoka T, Okuda K, Hirose T, Adachi M. [Three cases of *Mycobacterium kansasii* pulmonary diseases in previously healthy young women]. *Nihon Kokyuki Gakkai Zasshi.* 2011 Jun;49(6):426-31. Japanese.
- Otal I, Gomez AB, Kremer K, de Haas P, García MJ, Martín C, van Soolingen D. Mapping of IS6110 insertion sites in *Mycobacterium bovis* isolates in relation to adaptation from the animal to human host. *Vet Microbiol.* 2008 Jun 22;129(3-4):333-41. Epub 2007 Dec 8. Erratum in: *Vet Microbiol.* 2009 Mar 30;135(3-4):406. Isabel, Otal [corrected to Otal, Isabel].
- Pabst, M. J., J. M. Gross, J. P. Brozna, and M. B. Goren. 1988. Inhibition of macrophage priming by sulfatide from *Mycobacterium tuberculosis*.
- Panteix G, Gutierrez MC, Boschirololi ML, Rouviere M, Plaidy A, et al. (2010) Pulmonary tuberculosis

- due to *Mycobacterium microti*: a study of six recent cases in France. J Med Microbiol 59: 984–989.
- Petroff SA (1915) A New and Rapid Method for the Isolation and Cultivation of Tubercle 475 Bacilli Directly from the Sputum and Feces. J Exp Med 21: 38-42.
- Pfyffer GE, Auckenthaler R, van Embden JD, van Soolingen D. *Mycobacterium canettii*, the smooth variant of *M. tuberculosis*, isolated from a Swiss patient exposed in Africa. Emerg Infect Dis. 1998 Oct-Dec;4(4):631-4.
- Poitras E et Houde A. 2002. La PCR en temps réel: principes et applications. Reviews in Biology and Biotechnology Vol.2, No 2, December 2002. pp.2-11
- Radhakrishnan I, K MY, Kumar RA, Mundayoor S (2001) Implications of low frequency of IS6110 in fingerprinting field isolates of *Mycobacterium tuberculosis* from Kerala, India. J Clin Microbiol 39: 1683.
- Rajni, Rao N, Meena LS. Biosynthesis and Virulent Behavior of Lipids Produced by *Mycobacterium tuberculosis*: LAM and Cord Factor: An Overview. Biotechnol Res Int. 2011;2011:274693. Epub 2010 Dec 19.
- Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. Tuber Lung Dis. 1998;79(1):3-29.
- Rao, KR, F Kauser, S Srinivas, S Zanetti, LA Sechi, N Ahmed, SE Hasnain. 2005. Analysis of genomic downsizing on the basis of region-of-difference polymorphism profiling of *Mycobacterium tuberculosis* patient isolates reveals geographic partitioning. J Clin Microbiol. 43:5978-5982.
- Reed, G. 1957. Genus *Mycobacterium* (species affecting warm-blooded animals except those causing leprosy), In: R.S. BREED, E.G.D. MURRAY and N.R. SMITH (eds): Bergey's Manual of Determinative Bacteriology, 7th edition, The Williams & Wilkins Co, Baltimore, 1957, p. 703-704.
- Renfrew, C. 2010, Archaeogenetics--towards a 'new synthesis'? Curr 20:R162-165.
- Report Directorate of General of Health Services (2011) TB INDIA 2011 Revised 348 National TB Control Programme Annual Status Central TB Division, Government of 349 India.
- Ristow M, Möhlig M, Rifai M, Schatz H, Feldmann K, Pfeiffer A. New isoniazid/ethionamide resistance gene mutation and screening for multidrug-resistant *Mycobacterium tuberculosis* strains. Lancet. 1995 Aug 19;346(8973):502-3.
- Roberts, DE, D Ascherman, N Kleckner. 1991. IS10 promotes adjacent deletions at low frequency. Genetics. 128:37-43.
- Romagnani, S. 1996. Understanding the role of Th1/Th2 cells in infection. Trends Microbiol. 4:470–473.
- Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003 Aug 12;19(12):1572-4.
- Rouse DA, Li Z, Bai GH, Morris SL. Characterization of the katG and inhA genes of isoniazid-resistant clinical isolates of *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 1995a Nov;39(11):2472-7.
- Rouse DA, Morris SL. Molecular mechanisms of isoniazid resistance in *Mycobacterium tuberculosis* and *Mycobacterium bovis*. Infect Immun. 1995b Apr;63(4):1427-33.
- Sampson S, Warren R, Richardson M, van der Spuy G, van Helden P. IS6110 insertions in *Mycobacterium tuberculosis*: predominantly into coding regions. J Clin Microbiol. 2001 Sep;39(9):3423-4.
- Sampson, SL, RM Warren, M Richardson, TC Victor, AM Jordaan, GD van der Spuy, PD van Helden. 2003. IS6110-mediated deletion polymorphism in the direct repeat region of clinical isolates of *Mycobacterium tuberculosis*. J Bacteriol 185:2856-2866.
- Sherman DR, Mdluli K, Hickey MJ, Arain TM, Morris SL, Barry CE 3rd, Stover CK. Compensatory ahpC gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. Science. 1996 Jun 14;272(5268):1641-3.
- Schluger NW, Rom WN. The host immune response to tuberculosis. Am J Respir Crit Care Med. 1998 Mar;157(3 Pt 1):679-91. Review.
- Schurch, AC, K Kremer, A Kiers, MJ Boeree, RJ Siezen, D van Soolingen. 2011. Preferential deletion events in the direct repeat locus of *Mycobacterium tuberculosis*. J Clin Microbiol 49:1318-1322.
- Sharma P, Chauhan DS, Upadhyay P, Faujdar J, Lavania M, Sachan S, Katoch K, Katoch VM. Molecular typing of *Mycobacterium tuberculosis* isolates from a rural area of Kanpur by spoligotyping and mycobacterial interspersed repetitive units (MIRUs) typing. Infect Genet Evol. 2008 Sep;8(5):621-6. Epub 2008 May 9.
- Siddiqi N, Shamim M, Amin A, Chauhan DS, Das R, et al. (2001) Typing of drug resistant isolates of

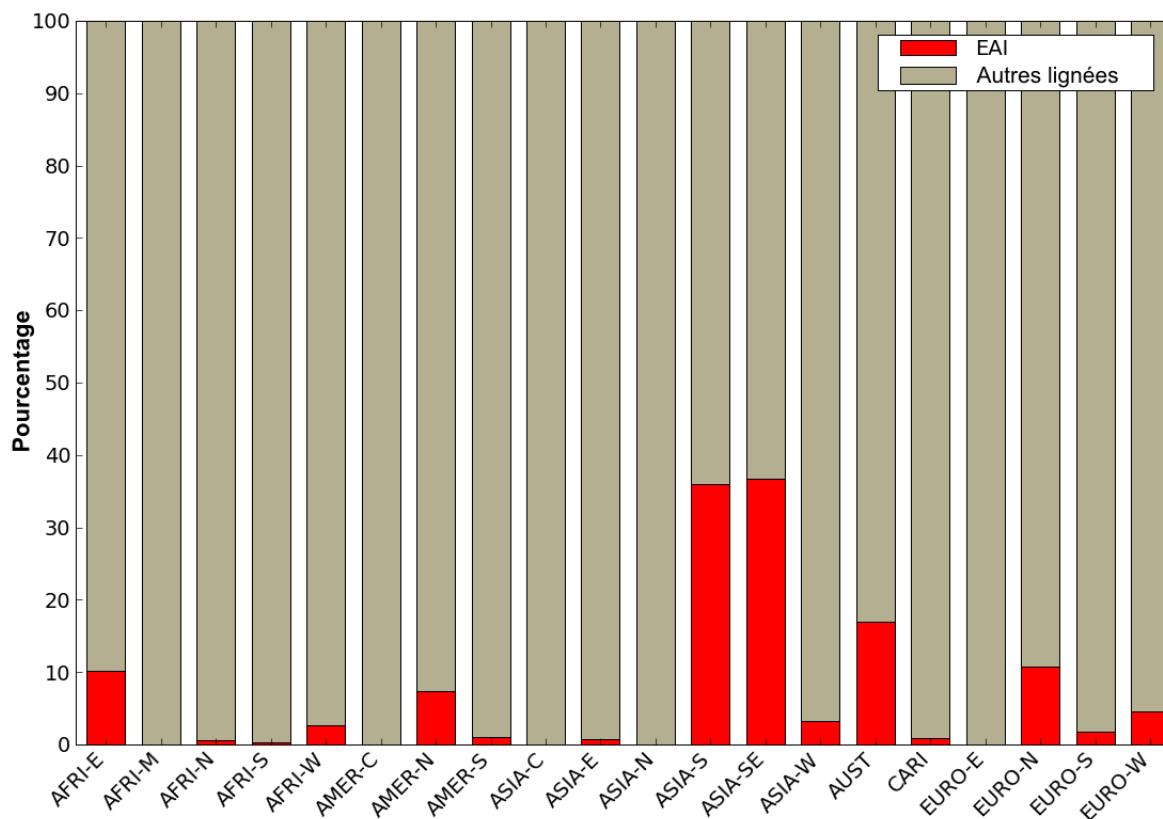
- Mycobacterium tuberculosis* from India using the IS6110 element reveals substantive polymorphism. *Infect Genet Evol* 1: 109-116.
- Simpson, E.H., 1949. Measurement of Diversity. *Nature*. 163: 688.
- Singh UB, Suresh N, Bhanu NV, Arora J, Pant H, et al. (2004) Predominant tuberculosis spoligotypes, Delhi, India. *Emerg Infect Dis* 10: 1138–1142.
- Slany M, Jezek P, Fiserova V, Bodnarova M, Stork J, Havelkova M, Kalat F, Pavlik I. *Mycobacterium marinum* infections in humans and tracing of its possible environmental sources. *Can J Microbiol*. 2012 Jan;58(1):39-44.
- Smith NH, Crawshaw T, Parry J, Birtles RJ (2009) *Mycobacterium microti*: More diverse than previously thought. *J Clin Microbiol* 47: 2551–2559.
- Smittipat N, Palittapongarnpim P. Identification of possible loci of variable number of tandem repeats in *Mycobacterium tuberculosis*. *Tuber Lung Dis*. 2000;80(2):69-74.
- Sola C, Devallois A, Horgen L, Maïsetti J, Filliol I, Legrand E, Rastogi N. Tuberculosis in the Caribbean: using spacer oligonucleotide typing to understand strain origin and transmission. *Emerg Infect Dis*. 1999 May-Jun;5(3):404-14.
- Sola C, Filliol I, Gutierrez MC, Mokrousov I, Vincent V, Rastogi N. Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. *Emerg Infect Dis*. 2001 May-Jun;7(3):390-6.
- Sola, C, S Ferdinand, LA Sechi, S Zanetti, D Martial, C Mammina, A Nastasi, G Fadda, N Rastogi. 2005. *Mycobacterium tuberculosis* molecular evolution in western Mediterranean Islands of Sicily and Sardinia. *Infect Genet Evol* 5:145-156.
- Sreevatsan S, Pan X, Zhang Y, Kreiswirth BN, Musser JM. Mutations associated with pyrazinamide resistance in *pncA* of *Mycobacterium tuberculosis* complex organisms. *Antimicrob Agents Chemother*. 1997a Mar;41(3):636-40,
- Sreevatsan S, Stockbauer KE, Pan X, Kreiswirth BN, Moghazeh SL, Jacobs WR Jr, Telenti A, Musser JM. Ethambutol resistance in *Mycobacterium tuberculosis*: critical role of *embB* mutations. *Antimicrob Agents Chemother*. 1997b Aug;41(8):1677-81.
- Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A*. 1997c Sep 2;94(18):9869-74.
- Stavrum R, Mphahlele M, Ovreås K, Muthivhi T, Fourie PB, Weyer K, Grewal HM. High diversity of *Mycobacterium tuberculosis* genotypes in South Africa and preponderance of mixed infections among ST53 isolates. *J Clin Microbiol*. 2009 Jun;47(6):1848-56. Epub 2009 Apr 22.
- Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsche-Gerdes S, Willery E, Savine E, de Haas P, van Deutekom H, Roring S, Bifani P, Kurepina N, Kreiswirth B, Sola C, Rastogi N, Vatin V, Gutierrez MC, Fauville M, Niemann S, Skuce R, Kremer K, Loch C, van Soolingen D. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol*. 2006 Dec;44(12):4498-510, Epub 2006 Sep 27. PubMed PMID: 17005759; PubMed Central PMCID: PMC1698431.
- Supply P, Lesjean S, Savine E, Kremer K, van Soolingen D, Loch C. Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. *J Clin Microbiol*. 2001 Oct;39(10):3563-71.
- Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Loch C. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol*. 2000 May;36(3):762-71.
- Supply P, Niemann S, Wirth T. On the mutation rates of spoligotypes and variable numbers of tandem repeat loci of *Mycobacterium tuberculosis*. *Infect Genet Evol*. 2011 Mar;11(2):251-2.
- Takiff HE, Salazar L, Guerrero C, Philipp W, Huang WM, Kreiswirth B, Cole ST, Jacobs WR Jr, Telenti A. Cloning and nucleotide sequence of *Mycobacterium tuberculosis* *gyrA* and *gyrB* genes and detection of quinolone resistance mutations. *Antimicrob Agents Chemother*. 1994 Apr;38(4):773-80,
- Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Colston MJ, Matter L, Schopfer K, Bodmer T. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet*. 1993 Mar 13;341(8846):647-50,
- Telenti A, Philipp WJ, Sreevatsan S, Bernasconi C, Stockbauer KE, Wieles B, Musser JM, Jacobs WR Jr. The *emb* operon, a gene cluster of *Mycobacterium tuberculosis* involved in resistance to ethambutol. *Nat Med*. 1997 May;3(5):567-70,
- Thierry D, Brisson-Noël A, Vincent-Lévy-Frébault V, Nguyen S, Guesdon JL, Gicquel B. Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS6110, and its application in diagnosis. *J Clin Microbiol*. 1990 Dec;28(12):2668-73.

- Thorne N, Underwood A, Gharbia S, Arnold C (2007) Evolutionary clues from comparative analysis of *Mycobacterium tuberculosis* variable-number tandem repeat sequences within genetic families. *Infect Genet Evol* 7: 239–246.
- Tibayrenc M, « *Encyclopedia of Infectious Diseases: Modern Methodologies* ». (New Jersey, USA : Wiley & Sons, Inc., Hoboken, 2007).
- Tsolaki AG, Gagneux S, Pym AS, Goguet de la Salmoniere YO, Kreiswirth BN, Van Soolingen D, Small PM. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J Clin Microbiol*. 2005 Jul;43(7):3185-91.
- UNAIDS. 2009. AIDS epidemic update, December 2009. WHO Library Cataloguing-in-Publication, Geneva, Switzerland. http://data.unaids.org/pub/Report/2009/jc1700_epi_update_2009_en.pdf.
- Underhill, PA, T Kivisild. 2007. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet*. 41:539-564.
- U'Ren, JM, JM Schupp, T Pearson, et al. 2007. Tandem repeat regions within the *Burkholderia pseudomallei* genome and their application for high resolution genotyping. *BMC Microbiol*. 7:23.
- van der Zanden, AG, K Kremer, LM Schouls, K Caimi, A Cataldi, A Hulleman, NJ Nagelkerke, D van Soolingen. 2002. Improvement of differentiation and interpretability of spoligotyping for *Mycobacterium tuberculosis* complex isolates by introduction of new spacer oligonucleotides. *J Clin Microbiol* 40:4628-4639.
- van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, Hermans P, Martin C, McAdam R, Shinnick TM, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol*. 1993 Feb;31(2):406-9.
- van Embden JD, van Gorkom T, Kremer K, Jansen R, van Der Zeijst BA, Schouls LM. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. *J Bacteriol*. 2000 May;182(9):2393-401.
- van Ingen J, Boeree MJ, van Soolingen D, Iseman MD, Heifets LB, Daley CL. Are phylogenetic position, virulence, drug susceptibility and in vivo response to treatment in mycobacteria interrelated? *Infect Genet Evol*. 2011 Oct 20,
- van Soolingen D, Hoogenboezem T, de Haas PE, Hermans PW, Koedam MA, Teppema KS, Brennan PJ, Besra GS, Portaels F, Top J, Schouls LM, van Embden JD. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. *Int J Syst Bacteriol*. 1997 Oct;47(4):1236-45.
- van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, Portaels F, Qing HZ, Enkhsaikan D, Nymadawa P, van Embden JD. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J Clin Microbiol*. 1995 Dec;33(12):3234-8.
- van Soolingen D, van der Zanden AG, de Haas PE, Noordhoek GT, Kiers A, Foudraïne NA, Portaels F, Kolk AH, Kremer K, van Embden JD. Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic markers. *J Clin Microbiol*. 1998 Jul;36(7):1840-5.
- Varma-Basil M, Kumar S, Arora J, Angrup A, Zozio T, Banavaliker JN, Singh UB, Rastogi N, Bose M. Comparison of spoligotyping, mycobacterial interspersed repetitive units typing and IS6110-RFLP in a study of genotypic diversity of *Mycobacterium tuberculosis* in Delhi, North India. *Mem Inst Oswaldo Cruz*. 2011 Aug;106(5):524-35.
- Vasconcellos SE, Huard RC, Niemann S, Kremer K, Santos AR, et al. (2010) Distinct genotypic profiles of the two major clades of *Mycobacterium africanum*. *BMC Infect Dis* 10: 80,
- Veziris N, Cambau E, Sougakoff W, Robert J, Jarlier V. [Resistance to antituberculous drugs]. *Arch Pediatr*. 2005 Aug;12 Suppl 2:S102-9. Review.
- Viegas SO, Machado A, Groenheit R, Ghebremichael S, Pennhag A, Gudo PS, Cuna Z, Miotto P, Hill V, Marrufo T, Cirillo DM, Rastogi N, Källenius G, Koivula T. 2010, Molecular diversity of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in Mozambique. *BMC Microbiol*. 2010 Jul 21;10:195.
- Wada T, Iwamoto T, Maeda S. Genetic diversity of the *Mycobacterium tuberculosis* Beijing family in East Asia revealed through refined population structure analysis. *FEMS Microbiol Lett*. 2009 Feb;291(1):35-43.
- Warren RM, Sampson SL, Richardson M, Van Der Spuy GD, Lombard CJ, Victor TC, van Helden PD. Mapping of IS6110 flanking regions in clinical isolates of *Mycobacterium tuberculosis* demonstrates genome plasticity. *Mol Microbiol*. 2000 Sep;37(6):1405-16.
- Warren RM, Streicher EM, Sampson SL, van der Spuy GD, Richardson M, et al. (2002) Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. *J Clin Microbiol* 40 : 4457–4465.
- Warren, R.M., Victor, T.C., Streicher, E.M., Richardson, M., Beyers, N., Gey van Pittius, N.C., van

- Helden, P.D., 2004. Patients with active tuberculosis often have different strains in the same sputum specimen. *Am. J. Respir. Crit. Care Med.* 169, 610–614.
- Wayne LG. International Committee on Systematic Bacteriology: announcement of the report of the ad hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Zentralbl Bakteriell Mikrobiol Hyg A.* 1988 Jun; 268(4):433-4.
- Wells AQ, Oxon DM. 1937. Tuberculosis in wild voles. *Lancet I* :1221.
- Wells AQ. *Mycobacterium tuberculosis* var. *muris*. *J Gen Microbiol.* 1953 Aug;9(1):149.
- Wirth T, Hildebrand F, Allix-Béguec C, Wölbeling F, Kubica T, Kremer K, van Soolingen D, Rüscher-Gerdes S, Locht C, Brisse S, Meyer A, Supply P, Niemann S. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog.* 2008 Sep 19;4(9):e1000160,
- World Health Organization, WHO 2009. Global tuberculosis control: epidemiology, strategy, financing: WHO, Geneva, Switzerland.
http://whqlibdoc.who.int/publications/2009/9789241598866_eng.pdf.
- World Health Organization, WHO Report 2011: Global Tuberculosis Control: WHO, Geneva, Switzerland.
- Yeh, R.W., Hopewell, P.C., Daley, C.L., 1999. Simultaneous infection with two strains of *Mycobacterium tuberculosis* identified by restriction fragment length polymorphism analysis. *Int. J. Tuberc. Lung Dis.* 3, 537–539.
- Yuen L.K.W., Ross B.C., Jackson K.M. and Dwyer B. ~ Characterization of *Mycobacterium tuberculosis* strains from Vietnamese patients by Southern blot hybridization. *J. Clin. Microbiol.*, 1993, 31:1615-1618.
- Zhang Y, Gladyshev VN. An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. *Bioinformatics.* 2005 Jun 1;21(11):2580-9.
- Zhang Y, Heym B, Allen B, Young D, Cole S. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature.* 1992 Aug 13;358(6387):591-3.
- Zhang Y. Genetic basis of isoniazid resistance of *Mycobacterium tuberculosis*. *Res Microbiol.* 1993 Feb;144(2):143-9. Review.
- Zhao Y, Feng Q, Tang K, Zhang C, Sun H, Luo T, Yang Z, Couvin D, Rastogi N, Sun Q. The population structure of drug-resistant *Mycobacterium tuberculosis* clinical isolates from Sichuan in China. *Infect Genet Evol.* 2011 Oct 2.
- Zimhony O, Vilchèze C, Arai M, Welch JT, Jacobs WR Jr. Pyrazinoic acid and its n-propyl ester inhibit fatty acid synthase type I in replicating tubercle bacilli. *Antimicrob Agents Chemother.* 2007 Feb;51(2):752-4. Epub 2006 Nov 13.
- Zozio T, Allix C, Gunal S, Saribas Z, Alp A, et al. (2005) Genotyping of *Mycobacterium tuberculosis* clinical isolates in two cities of Turkey: description of a new family of genotypes that is phylogeographically specific for Asia Minor. *BMC Microbiol* 5: 44.
- Zozio T, « *Diversité des bacilles tuberculeux circulant sur la planète* ». (Allemagne : Editions Universitaires Européennes, 2011).

1081	2	Inconnue	2	-	-
1915	2	S	2	-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T4		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	H3		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T3		-	-
	2	T1		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T2		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T3		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	LAM9		-	-
	2	Inconnue		-	-
	2	T3		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T1		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T5		-	-
	2	Inconnue		-	-
	2	T1		-	-
	2	Inconnue		-	-
	2	T4		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T1		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T5		-	-
	2	Inconnue		-	-
	2	Inconnue		-	-
	2	T1		-	-

Annexe 4. Distribution des EAI dans les sous-régions géographiques du monde, (données tirées de SpoIDB4)



Annexe 5. Reclassification de 176 profils de la base de données en ligne MIRU-VNTR^{plus} (<http://www.miru-vntrplus.org/MIRU/index.faces>).

ID	Lignée MIRU12 prédite par VNTR-Tbbayes	Probabilité de VNTR-Tbbayes	Lignée MIRU12 par la méthode CZS	Lignée	Lignée LSP	MIRU12	MIT
7936/01	East Asian-101	0,84	East Asian-101	Dehli/CAS	East African Indian	225226173523	
7746/01	East Asian-93	0,67	East Asian-93	Dehli/CAS	East African Indian	226225173533	932
1805/02	East Asian-93	0,67	East-African Indian-68	Dehli/CAS	East African Indian	224425133533	
2637/02	East-African Indian-68	0,79	East Asian-101	Dehli/CAS	East African Indian	225526173523	
6427/01	East-African Indian-68	0,84	East-African Indian-68	Dehli/CAS	East African Indian	225426173523	
7507/01	East-African Indian-68	0,8	East-African Indian-68	Dehli/CAS	East African Indian	226425193433	
7747/01	East-African Indian-68	0,92	East-African Indian-68	Dehli/CAS	East African Indian	225426143523	
8260/01	East-African Indian-68	0,98	East-African Indian-68	Dehli/CAS	East African Indian	225424183433	
9398/01	East-African Indian-68	0,84	East-African Indian-68	Dehli/CAS	East African Indian	225426173523	
9915/01	East-African Indian-68	0,84	East-African Indian-68	Dehli/CAS	East African Indian	225426173523	
3256/02	East Asian-16	1	East Asian-16	Beijing	East Asian	223325153533	16
3277/02	East Asian-16	1	East Asian-16	Beijing	East Asian	223325153533	16
3329/02	East Asian-16	1	East Asian-16	Beijing	East Asian	223325153533	16
3364/02	East Asian-16	1	East Asian-16	Beijing	East Asian	223325153533	16
4445/02	East Asian-16	1	East Asian-16	Beijing	East Asian	223325153533	16
4498/02	East Asian-16	1	East Asian-16	Beijing	East Asian	223325153533	16
4499/02	East Asian-16	1	East Asian-16	Beijing	East Asian	223325153533	16
3243/02	East Asian-17	1	East Asian-17	Beijing	East Asian	223325173533	17
4436/02	East Asian-17	1	East Asian-17	Beijing	East Asian	223325173533	17
3309/02	East Asian-99	0,64	East Asian-99	Beijing	East Asian	223325153543	1153
10493/01	East Asian-86	0,81	East Asian-86	Ghana	Euro-American	223325143436	
2253/99	Euro American-112	0,77	Euro American-112	Ugandall	Euro-American	223315143424	821
1571/99	Euro American-116	0,83	Euro American-116	Ugandal	Euro-American	223225153423	1264
2111/99	Euro American-116	0,83	Euro American-116	Ugandal	Euro-American	223225153423	1264
10438/01	Euro American-12	0,99	Euro American-12	Cameroon	Euro-American	223215153323	264
10439/01	Euro American-12	0,97	Euro American-12	Cameroon	Euro-American	223315153323	12
10445/01	Euro American-12	0,97	Euro American-12	Cameroon	Euro-American	223315153323	12
10446/01	Euro American-12	0,62	Euro American-12	Cameroon	Euro-American	123215143323	
10481/01	Euro American-12	0,97	Euro American-12	Cameroon	Euro-American	223315153323	12
1417/02	Euro American-12	0,97	Euro American-12	Cameroon	Euro-American	223315153323	12
1428/02	Euro American-12	0,97	Euro American-12	Cameroon	Euro-American	223315153323	12
5390/02	Euro American-12	0,99	Euro American-12	Cameroon	Euro-American	223215153323	264
5429/02	Euro American-12	0,97	Euro American-12	Cameroon	Euro-American	223315153323	12
1647/99	Euro American-12	0,58	Euro American-12	Ugandall	Euro-American	223315143323	784
2176/99	Euro American-12	0,89	Euro American-12	Ugandall	Euro-American	223315143321	1144
10469/01	Euro American-121	0,9	Euro American-121	Ghana	Euro-American	223325143335	
10470/01	Euro American-121	0,85	Euro American-121	Ghana	Euro-American	223325143336	673
10486/01	Euro American-121	0,85	Euro American-121	Ghana	Euro-American	223325143336	673
10515/01	Euro American-121	0,85	Euro American-121	Ghana	Euro-American	223325143336	673
1438/02	Euro American-121	0,9	Euro American-121	Ghana	Euro-American	223325143335	
2570/02	Euro American-121	0,9	Euro American-121	Ghana	Euro-American	223325143335	
2582/02	Euro American-121	0,9	Euro American-121	Ghana	Euro-American	223325143335	
2597/02	Euro American-121	0,9	Euro American-121	Ghana	Euro-American	223325143335	
5357/02	Euro American-121	0,77	Euro American-121	Ghana	Euro-American	223325143337	
2329/99	Euro American-121	0,46	East Asian-101	Ugandal	Euro-American	223325133223	
2263/99	Euro American-121	0,85	Euro American-121	Ugandal	Euro-American	223225143423	
1521/99	Euro American-121	0,68	Euro American-12	Ugandall	Euro-American	223315144423	
2307/99	Euro American-121	0,68	Euro American-12	Ugandall	Euro-American	223315144423	
2191/99	Euro American-121	0,67	Euro American-121	Ugandall	Euro-American	223315143423	11
2197/99	Euro American-121	0,67	Euro American-121	Ugandall	Euro-American	223315143423	11
2211/99	Euro American-121	0,67	Euro American-121	Ugandall	Euro-American	223315143423	11
2319/99	Euro American-121	0,67	Euro American-121	Ugandall	Euro-American	223315143423	11
2379/99	Euro American-121	0,67	Euro American-121	Ugandall	Euro-American	223315143423	11
8885/03	Euro American-128	0,82	Euro American-128	LAM	Euro-American	223226141321	
946/03	Euro American-128	0,65	Euro American-128	LAM	Euro-American	223126153321	536
3686/03	Euro American-15	0,67	Euro American-15	Haarlem	Euro-American	224325153322	32
4217/02	Euro American-15	0,79	Euro American-15	Haarlem	Euro-American	224322153322	
6946/03	Euro American-15	0,79	Euro American-15	Haarlem	Euro-American	224322153322	
2169/99	Euro American-15	0,56	Euro American-15	Ugandal	Euro-American	223225143222	
2173/99	Euro American-15	0,42	Euro American-15	Ugandal	Euro-American	223225143422	
2201/99	Euro American-15	0,99	Euro American-15	Ugandal	Euro-American	223325153422	
2224/99	Euro American-15	0,93	Euro American-15	Ugandal	Euro-American	223325153522	336
2331/99	Euro American-15	0,97	Euro American-15	Ugandal	Euro-American	223225153422	1007
2333/99	Euro American-15	0,97	Euro American-15	Ugandal	Euro-American	223225153422	1007
10581/03	Euro American-190	0,99	Euro American-190	LAM	Euro-American	124325153224	326
3262/02	Euro American-190	0,99	Euro American-190	LAM	Euro-American	124325153224	326
3310/02	Euro American-190	0,99	Euro American-190	LAM	Euro-American	124325143225	325
4431/02	Euro American-190	0,99	Euro American-190	LAM	Euro-American	124325143225	325
8078/03	Euro American-190	0,99	Euro American-190	LAM	Euro-American	124326153224	140
4428/02	Euro American-190	0,51	Euro American-213	LAM	Euro-American	224327153226	
11046/04	Euro American-212	1	Euro American-212	S	Euro-American	333325153324	397
1897/04	Euro American-212	0,97	Euro American-212	S	Euro-American	333325153424	

2151/03	Euro American-212	0,99	Euro American-212	S	Euro-American	233325153324	212
2318/06	Euro American-212	0,96	Euro American-212	S	Euro-American	332325153325	
282/04	Euro American-212	0,88	Euro American-212	S	Euro-American	233325133326	
3270/04	Euro American-212	0,92	Euro American-212	S	Euro-American	233225153326	
4526/04	Euro American-212	0,99	Euro American-212	S	Euro-American	333225143325	216
6411/05	Euro American-212	0,92	Euro American-212	S	Euro-American	333325154325	
6424/05	Euro American-212	1	Euro American-212	S	Euro-American	333325143324	759
742/06	Euro American-212	0,96	Euro American-212	S	Euro-American	332325153325	
7955/03	Euro American-212	0,95	Euro American-212	S	Euro-American	233225153324	
8583/04	Euro American-212	0,85	Euro American-212	S	Euro-American	333325134222	
3995/03	Euro American-213	0,97	Euro American-213	LAM	Euro-American	224326153324	215
7968/03	Euro American-213	0,99	Euro American-213	LAM	Euro-American	224326153323	213
9953/04	Euro American-213	0,87	Euro American-213	X	Euro-American	224326153322	1062
1850/03	Euro American-25	0,99	Euro American-25	LAM	Euro-American	224226143321	738
5400/02	Euro American-33	0,79	Euro American-33	Cameroon	Euro-American	224315153323	27
4412/04	Euro American-33	0,67	Euro American-33	X	Euro-American	224325153325	35
9787/04	Euro American-33	0,67	Euro American-33	X	Euro-American	224325153325	35
10264/03	Euro American-40	1	Euro American-40	TUR	Euro-American	215125113322	310
10529/03	Euro American-40	1	Euro American-40	TUR	Euro-American	215125113322	310
11313/03	Euro American-40	1	Euro American-40	TUR	Euro-American	215125113322	310
2258/03	Euro American-40	1	Euro American-40	TUR	Euro-American	215125113322	310
8577/03	Euro American-40	0,53	East Asian-101	URAL	Euro-American	227225113223	171
1657/03	Euro American-40	1	Euro American-40	URAL	Euro-American	229225113322	
2679/03	Euro American-40	0,98	Euro American-40	URAL	Euro-American	239225113322	
12637/02	Euro American-43	0,99	Euro American-43	Haarlem	Euro-American	225323153323	43
3103/03	Euro American-43	0,99	Euro American-43	Haarlem	Euro-American	225323153323	43
2336/02	Euro American-45	0,99	Euro American-45	Haarlem	Euro-American	225325153323	45
3342/02	Euro American-45	0,92	Euro American-45	Haarlem	Euro-American	225325151323	762
4130/02	Euro American-45	0,99	Euro American-45	Haarlem	Euro-American	225325153323	45
4993/02	Euro American-45	0,99	Euro American-45	Haarlem	Euro-American	225325153323	45
9532/03	Euro American-45	0,99	Euro American-45	Haarlem	Euro-American	225325153323	45
9400/02	Euro American-46	0,65	Euro American-46	Haarlem	Euro-American	225324153324	
8750/03	Euro American-7	0,37	Euro American-15	Haarlem	Euro-American	226325153322	1094
4192/03	Euro American-7	0,95	Euro American-7	Haarlem	Euro-American	222325153323	7
8870/03	Euro American-7	0,69	Euro American-121	NEW-1	Euro-American	222225143323	547
10459/03	Euro American-7	0,95	Euro American-7	NEW-1	Euro-American	222325153323	7
12591/02	Euro American-7	0,96	Euro American-7	NEW-1	Euro-American	222325163323	111
8915/03	Indo-Oceanic-56	0,99	Indo-Oceanic-56	EAI	Indo-Oceanic	254326223432	56
9267/01	Indo-Oceanic-56	0,99	Indo-Oceanic-56	EAI	Indo-Oceanic	254326223432	56
947/01	Indo-Oceanic-57	0,95	Indo-Oceanic-56	EAI	Indo-Oceanic	244327221632	
11359/03	Indo-Oceanic-57	0,97	Indo-Oceanic-57	EAI	Indo-Oceanic	255316223534	
6538/03	Indo-Oceanic-57	0,99	Indo-Oceanic-57	EAI	Indo-Oceanic	254326223533	57
7190/03	Indo-Oceanic-57	1	Indo-Oceanic-57	EAI	Indo-Oceanic	254426223533	1003
11051/03	Indo-Oceanic-59	0,82	Indo-Oceanic-59	EAI	Indo-Oceanic	244225223533	
12778/03	Indo-Oceanic-59	1	Indo-Oceanic-59	EAI	Indo-Oceanic	264225223533	59
1797/03	Indo-Oceanic-64	0,67	Indo-Oceanic-64	EAI	Indo-Oceanic	254326223643	
6006/03	Indo-Oceanic-64	0,99	Indo-Oceanic-64	EAI	Indo-Oceanic	244326123413	
4058/03	Indo-Oceanic-69	0,97	Indo-Oceanic-69	EAI	Indo-Oceanic	254324223434	
4850/03	Indo-Oceanic-69	0,98	Indo-Oceanic-69	EAI	Indo-Oceanic	253327223334	
1290/03	M. bovis	0,98	M. bovis	Bovis	M. bovis	232126213322	
1601/01	M. bovis	1	M. bovis	Bovis	M. bovis	232324253322	49
4258/00	M. bovis	1	M. bovis	Bovis	M. bovis	232324253222	
5346/02	M. bovis	1	M. bovis	Bovis	M. bovis	222324253322	5
7072/01	M. bovis	1	M. bovis	Bovis	M. bovis	232324263322	
751/01	M. bovis	1	M. bovis	Bovis	M. bovis	232324263322	
7540/01	M. bovis	1	M. bovis	Bovis	M. bovis	232324252322	1321
8217/02	M. bovis	1	M. bovis	Bovis	M. bovis	232324243322	
8490/00	M. bovis	0,8	M. bovis	Bovis	M. bovis	242424133322	
951/01	M. bovis	1	M. bovis	Bovis	M. bovis	252324253322	
9564/00	M. bovis	0,99	M. bovis	Bovis	M. bovis	232224263222	
3040/99	M. canettii	0,98	M.canettii	Canetti	M. canetti	313318672636	
3041/99	M. canettii	1	M.canettii	Canetti	M. canetti	323212632428	60
11443/99	West African II	1	West African lineage II	Caprae	M. caprae	235424253522	
1694/00	West African II	1	West African lineage II	Caprae	M. caprae	225524253522	
1696/00	West African II	1	West African lineage II	Caprae	M. caprae	136324243522	
5358/99	West African II	1	West African lineage II	Caprae	M. caprae	236224243522	
7140/99	West African II	1	West African lineage II	Caprae	M. caprae	225424253522	
7618/99	West African II	1	West African lineage II	Caprae	M. caprae	246424253622	
8319/99	West African II	1	West African lineage II	Caprae	M. caprae	237324243522	
8522/00	West African II	1	West African lineage II	Caprae	M. caprae	246424323522	
8986/99	West African II	0,87	West African lineage II	Caprae	M. caprae	246324253422	
9062/01	West African II	0,99	West African lineage II	Caprae	M. caprae	234323243522	672
9577/99	West African II	1	West African lineage II	Caprae	M. caprae	236424253522	664
10400/02	West African I	1	West African lineage I	West African 1	West African-1	224324244221	1282
10458/02	West African I	1	West African lineage I	West African 1	West African-1	224424243221	348
10473/01	West African I	1	West African lineage I	West African 1	West African-1	224424243221	348
10480/01	West African I	1	West African lineage I	West African 1	West African-1	224424243221	348
10494/01	West African I	1	West African lineage I	West African 1	West African-1	227524242221	
1410/02	West African I	1	West African lineage I	West African 1	West African-1	224424244221	934
5473/02	West African I	1	West African lineage I	West African 1	West African-1	223424244221	

8303/02	West African I	1	West African lineage I	West African 1	West African-1	224424243221	348
4804/03	West African I	1	West African lineage I	West African 1	West African-1	224424243221	348
5398/02	West African I	1	West African lineage I	West African 1	West African-1	224424244221	934
5432/02	West African I	1	West African lineage I	West African 1	West African-1	224424243221	348
5434/02	West African I	1	West African lineage I	West African 1	West African-1	223424244221	
1443/02	West African I	1	West African lineage I	West African 1	West African-1	227524242221	
1449/02	West African I	1	West African lineage I	West African 1	West African-1	224424244221	934
1465/02	West African I	1	West African lineage I	West African 1	West African-1	214324244221	
1473/02	West African I	1	West African lineage I	West African 1	West African-1	223524244221	
2569/02	West African I	1	West African lineage I	West African 1	West African-1	224424243221	348
2577/02	West African I	1	West African lineage I	West African 1	West African-1	234424244221	
3482/03	West African I	1	West African lineage I	West African 1	West African-1	224424242221	
4802/03	West African I	1	West African lineage I	West African 1	West African-1	224624244221	
5468/02	West African I	0,69	West African lineage I	West African 2	West African-2	227424243521	
8236/02	West African I	0,58	West African lineage I	West African 2	West African-2	224424243622	
10514/01	West African II	0,81	West African lineage I	West African 2	West African-2	225422244522	
10462/01	West African II	1	West African lineage II	West African 2	West African-2	226424243522	
10476/01	West African II	1	West African lineage II	West African 2	West African-2	225424243522	304
10485/01	West African II	1	West African lineage II	West African 2	West African-2	225424243522	304
10512/01	West African II	1	West African lineage II	West African 2	West African-2	225424243522	304
10517/01	West African II	0,99	West African lineage II	West African 2	West African-2	225324243522	
5383/02	West African II	1	West African lineage II	West African 2	West African-2	225424243522	304
8163/02	West African II	0,98	West African lineage II	West African 2	West African-2	227424243522	
9550/00	West African II	0,98	West African lineage II-M.	West African 2	West African-2	237224243421	

Annexe 6. Reclassification de 190 souches publiées (Lazzarini et al. 2007). Cette collection rassemble 93 souches de Rio de Janeiro avec un RD(Rio) déléte, 7 souches internationales sans RD(Rio), et 90 autres provenant de Rio de Janeiro qui ont la région RD(Rio).

Souches	MIRU12	classification MIRU-VNTR
<u>Souches sans RD(Rio)</u>		
3(c)	224226153321	Euro American-25
10	224226143321	Euro American-25
14	224226153321	Euro American-25
17	224226133321	Euro American-25
25	224224164321	Euro American-25
32(a)	224226143321	Euro American-25
34	224226143321	Euro American-25
35(a)	225126163311	Euro American-25
37	224216151321	Euro American-25
38(a)	224224133321	Euro American-25
39	224226153321	Euro American-25
40(b)	224226143321	Euro American-25
41(a)	224226143321	Euro American-25
43	224226153321	Euro American-25
44	234225153321	Euro American-25
45	224224133321	Euro American-25
47(b)	224225153321	Euro American-25
48	224225153321	Euro American-25
57	224225163321	Euro American-25
61(c)	224226143321	Euro American-25
67	224226153321	Euro American-25
72	224222163321	Euro American-25
74	224226133321	Euro American-25
75	224226133321	Euro American-25
78	224224133321	Euro American-25
82(b)	224125133321	Euro American-25
84	224226153321	Euro American-25
87	224216151321	Euro American-25
88	224225153321	Euro American-25
91	224225133221	Euro American-25
92	224225143321	Euro American-25
93	224226153321	Euro American-25
42	224115153321	Euro American-34
50	224325143221	Euro American-121
2	223126133321	Euro American-128
6	223226143321	Euro American-128
7	223226152331	Euro American-128
18	222226143321	Euro American-128
22	223226143321	Euro American-128
23	223226153321	Euro American-128
26	223226123321	Euro American-128
27	223224153321	Euro American-128
28	223224153321	Euro American-128
29(b)	223226143321	Euro American-128
29(c)	223226143321	Euro American-128
30	223226163321	Euro American-128
31	223226153321	Euro American-128
33	222226163321	Euro American-128
36	223226163321	Euro American-128

51	223226153321	Euro American-128
54(b)	223124143321	Euro American-128
71	223226153321	Euro American-128
73	223235153321	Euro American-128
77(b)	223226153321	Euro American-128
85(b)	222226153321	Euro American-128
94	222225163321	Euro American-128
1	223125132221	Euro American-163
4	224126152321	Euro American-163
5	223126152311	Euro American-163
8	224126152321	Euro American-163
9	224124132321	Euro American-163
11(a)	222126152321	Euro American-163
13(a)	223126142221	Euro American-163
15	224126142321	Euro American-163
16(b)	224126142321	Euro American-163
19(a)	222226142321	Euro American-163
20	223126142321	Euro American-163
24	225226162321	Euro American-163
46	223126142321	Euro American-163
49(b)	224226152321	Euro American-163
52	224214132321	Euro American-163
53	224225152221	Euro American-163
55	222124142321	Euro American-163
56	223126152321	Euro American-163
58	224126151321	Euro American-163
63	223126152311	Euro American-163
64(c)	222124142321	Euro American-163
65	222124162321	Euro American-163
66(a)	222126152321	Euro American-163
68	224226152321	Euro American-163
76	224224132221	Euro American-163
79	224126152321	Euro American-163
80	224126162321	Euro American-163
81	224126152321	Euro American-163
86	222124142321	Euro American-163
89	222126152321	Euro American-163
90	223126142321	Euro American-163
83(b)	224326163321	Euro American-213
59	225326143221	Euro American-224
60	225326163221	Euro American-224
62	223127151321	Euro American-224
69	124326153321	Euro American-246
70	124326153321	Euro American-246

Souches internationales sans RD(Rio)

Percy47 (Djibouti)	224226153321	Euro American-25
14395 (AZ, USA)	224226153321	Euro American-25
15376 (AZ, USA)	224226153321	Euro American-25
2002-1330 (Netherlands)	224226133321	Euro American-25
2002-1384 (Netherlands)	224226133321	Euro American-25
00-817 (IN, USA)	223226153321	Euro American-128
94-1055 (Netherlands)	223124142321	Euro American-163

Souches sauvages - RD(Rio) présent

144	223115152324	Euro American-8
-----	--------------	-----------------

100	224226153323	Euro American-25
102	224226153323	Euro American-25
109(a)	224226153323	Euro American-25
134	124225133323	Euro American-33
149	224324153124	Euro American-34
163	225213152423	Euro American-42
110	225226153324	Euro American-46
103	223322153323	Euro American-116
112(a)	223226153323	Euro American-128
151	223216143323	Euro American-128
95	124226143224	Euro American-190
96	124326153222	Euro American-190
99	114226153224	Euro American-190
116	123226133225	Euro American-190
117	124326153225	Euro American-190
118	124226153227	Euro American-190
121	124326153222	Euro American-190
122	124326153224	Euro American-190
123	124326152222	Euro American-190
125	125326162227	Euro American-190
126(b)	123226153224	Euro American-190
127(a)	122325133224	Euro American-190
127(b)	122325133224	Euro American-190
128	124326153222	Euro American-190
130(a)	122326153224	Euro American-190
131	123326162224	Euro American-190
132	122326143227	Euro American-190
135(a)	124226162227	Euro American-190
136	123326162227	Euro American-190
137	123326162227	Euro American-190
138(a)	124326143222	Euro American-190
138(b)	124326143222	Euro American-190
139	124226143222	Euro American-190
140(b)	124326143222	Euro American-190
141	124226151223	Euro American-190
142	123327143226	Euro American-190
143(a)	125326153224	Euro American-190
145	123326133224	Euro American-190
146	123326133223	Euro American-190
147	124326152221	Euro American-190
152	124316143222	Euro American-190
153(b)	124226143224	Euro American-190
159	122326153225	Euro American-190
160	124326143222	Euro American-190
161	122326153225	Euro American-190
162	124326153224	Euro American-190
167(a)	124426154222	Euro American-190
167(b)	124426154222	Euro American-190
168	124326153222	Euro American-190
173	123326153222	Euro American-190
174	123326153222	Euro American-190
175	123326153221	Euro American-190
176	123326153222	Euro American-190
177(a)	123326152232	Euro American-190

178	123326143222	Euro American-190
179	224316153228	Euro American-190
180	224316143229	Euro American-190
158	133125161225	Euro American-212
164	333225162322	Euro American-212
165	333225152222	Euro American-212
101	224326153323	Euro American-213
105	224326153323	Euro American-213
106(b)	224326153323	Euro American-213
107(a)	224326153323	Euro American-213
111	224326153323	Euro American-213
113(b)	224326153323	Euro American-213
115	224326153324	Euro American-213
150	224326133324	Euro American-213
157	224316133325	Euro American-213
166	224326153324	Euro American-213
169(b)	224326133322	Euro American-213
170(a)	222326133323	Euro American-213
171	224326133323	Euro American-213
172(a)	223326133323	Euro American-213
181	224326153125	Euro American-213
104	213326153324	Euro American-224
108(a)	223326153324	Euro American-224
148	223316153127	Euro American-224
154	223326153324	Euro American-224
155	223326153324	Euro American-224
156	223326153324	Euro American-224
97(b)	124226153325	Euro American-246
98	124326153324	Euro American-246
114	124326143324	Euro American-246
119	124326133325	Euro American-246
120(a)	124325133326	Euro American-246
124	124326153323	Euro American-246
129	124325133322	Euro American-246
133	124326153325	Euro American-246

Annexe 7. Distribution des lignées prédites par VNTR-TBbayes à l'intérieur des lignées assignées avec la méthode du Z-Score Cumulé (CZS). Exemple : parmi les souches classifiées Indo-Oceanic-57 par la méthode CZS (= cluster 57), 99% et 1% sont respectivement prédites Indo-Oceanic-57 et Indo-Oceanic-59 par VNTR-Tbbaies. Ce tableau fut réalisé à partir de 2302 patterns MIRU-VNTR de SITVITWEB.

<u>Indo-Oceanic</u>			
<u>Cluster 57</u>		<u>Cluster 64</u>	
Indo-Oceanic-57	99	Indo-Oceanic-64	89,71
Indo-Oceanic-59	1	Indo-Oceanic-57	10,29
<u>Cluster 56</u>		<u>Cluster 69</u>	
Indo-Oceanic-56	94,12	Indo-Oceanic-69	96,55
Indo-Oceanic-57	5,88	Indo-Oceanic-57	1,72
<u>Cluster 59</u>		Indo-Oceanic-64	
Indo-Oceanic-59	93,33		1,72
Indo-Oceanic-57	4,44		
West African I	2,22		
<u>East Asian and African Indian</u>			
<u>Cluster 17</u>		<u>Cluster 99</u>	
East Asian-17	89,29	East Asian-99	93,55
East-African Indian-68	3,57	Euro American-116	6,45
East Asian-86	3,57	<u>Cluster 101</u>	
East Asian-83	1,79	East Asian-101	79,71
Indo-Oceanic-57	1,79	East Asian-93	5,80
<u>Cluster 16</u>		East Asian-17	4,35
East Asian-16	84,85	East Asian-86	2,90
East-African Indian-68	12,12	Euro American-121	1,45
East Asian-17	3,03	Euro American-15	1,45
<u>Cluster 83</u>		Euro American-213	1,45
East Asian-83	84	Euro American-224	1,45
East Asian-93	8	Euro American-40	1,45
East-African Indian-68	4	<u>Cluster 68</u>	
Euro American-8	4	East-African Indian-68	89,93
<u>Cluster 86</u>		East Asian-93	4,32
East Asian-86	78,18	East-African Indian-261	1,44
East Asian-17	10,91	East Asian-101	1,44
East-African Indian-68	3,64	East Asian-86	1,44
East-African Indian-261	1,82	Indo-Oceanic-59	1,44
East Asian-93	1,82	<u>Cluster : 261</u>	
Euro American-34	1,82	East-African Indian-261	60,98
Indo-Oceanic-56	1,82	East-African Indian-68	21,95
<u>Cluster 93</u>		East Asian-86	4,88
East Asian-93	78,69	Euro American-40	4,88
East-African Indian-68	14,75	East Asian-101	2,44
East Asian-17	4,92	East Asian-93	2,44
East Asian-16	1,64	Euro American-46	2,44
<u>West African I</u>			
<u>Cluster 934</u>			
West African I	83,33		
West African II	16,67		
<u>West African II</u>			
<u>Cluster 664</u>			
West African II	97,01		
Indo-Oceanic-57	1,49		
West African I	1,49		
<u>M. bovis</u>			
<u>Cluster 49</u>			
<i>M. bovis</i>	92,86		
West African II	4,76		
Indo-Oceanic-56	2,38		
<u>M. canettii</u>			
<u>Cluster 60</u>			
<i>M. canettii</i>	100		

Euro American

<u>Cluster : 116</u>		<u>Cluster : 34</u>	
Euro American-116	67,44	Euro American-34	79,59
Euro American-224	9,30	Euro American-33	8,16
Euro American-125	6,98	Euro American-15	3,06
East Asian-101	4,65	Euro American-190	2,04
Euro American-8	4,65	East Asian-93	1,02
Euro American-121	2,33	Euro American-125	1,02
Euro American-212	2,33	Euro American-212	1,02
Euro American-45	2,33	Euro American-213	1,02
<u>Cluster : 7</u>		Euro American-25	1,02
Euro American-7	78,41	Indo-Oceanic-56	1,02
Euro American-125	3,41	M. canettii	1,02
East Asian-101	2,27	<u>Cluster : 40</u>	
East Asian-99	2,27	Euro American-40	98
Euro American-116	2,27	Euro American-7	2
Euro American-121	2,27	<u>Cluster : 46</u>	
Euro American-46	2,27	Euro American-46	87,18
Euro American-8	2,27	Euro American-112	2,56
Euro American-15	1,14	Euro American-190	2,56
Euro American-40	1,14	Euro American-34	2,56
Euro American-45	1,14	Euro American-42	2,56
M. bovis	1,14	Euro American-43	2,56
<u>Cluster : 8</u>		<u>Cluster : 112</u>	
Euro American-8	85,14	Euro American-112	70,31
Euro American-112	6,76	Euro American-121	4,69
Euro American-212	2,70	Euro American-213	4,69
Euro American-12	1,35	Euro American-125	3,13
Euro American-15	1,35	Euro American-15	3,13
Euro American-46	1,35	Euro American-34	3,13
Euro American-7	1,35	East Asian-101	1,56
<u>Cluster : 12</u>		Euro American-12	1,56
Euro American-12	74,29	Euro American-190	1,56
Euro American-112	5,71	Euro American-212	1,56
Euro American-121	5,71	Euro American-224	1,56
Euro American-40	5,71	Euro American-246	1,56
Euro American-15	2,86	Euro American-8	1,56
Euro American-34	2,86	<u>Cluster : 121</u>	
Euro American-42	2,86	Euro American-121	72,60
<u>Cluster : 15</u>		Euro American-7	8,22
Euro American-15	88,89	Euro American-40	5,48
Euro American-7	3,03	Euro American-213	4,11
Euro American-121	1,01	Euro American-34	4,11
Euro American-125	1,01	Euro American-15	2,74
Euro American-212	1,01	East Asian-101	1,37
Euro American-224	1,01	Euro American-212	1,37
Euro American-33	1,01	<u>Cluster : 125</u>	
Euro American-34	1,01	Euro American-125	77,78
Euro American-46	1,01	East Asian-99	3,70
West African II	1,01	Euro American-12	3,70
<u>Cluster : 33</u>		Euro American-190	3,70
Euro American-33	77,50	Euro American-213	3,70
Euro American-121	7,50	Euro American-224	3,70
Euro American-15	5	Euro American-46	3,70
East Asian-101	2,50	<u>Cluster : 212</u>	
Euro American-213	2,50	Euro American-212	93,48
Euro American-34	2,50	Euro American-33	2,17
Euro American-40	2,50	Euro American-116	1,09
		Euro American-121	1,09
		Euro American-128	1,09
		Euro American-213	1,09

Euro American

LAM-25/128/163/224 (RD(Rio) deleted)

<u>Cluster : 25</u>		<u>Cluster : 163</u>	
Euro American-25	83,56	Euro American-163	93,75
Euro American-128	2,74	Euro American-25	6,25
Euro American-15	2,74	<u>Cluster : 224</u>	
Euro American-213	2,74	Euro American-224	92,16
Euro American-246	2,74	Euro American-7	3,92
Euro American-212	1,37	Euro American-163	1,96
Euro American-34	1,37	Euro American-213	1,96
Euro American-40	1,37		
Euro American-7	1,37		

Cluster : 128

Euro American-128	92,86
Euro American-12	2,38
Euro American-163	2,38
Euro American-212	2,38

LAM-190/213/246 (RD(Rio) WT)

<u>Cluster : 190</u>		<u>Cluster : 246</u>	
Euro American-190	100	Euro American-246	97,62
<u>Cluster : 213</u>		Euro American-34	2,38
Euro American-213	74,29		
Euro American-190	7,14		
Euro American-224	4,29		
Euro American-25	2,86		
Euro American-33	2,86		
East Asian-101	1,43		
Euro American-116	1,43		
Euro American-15	1,43		
Euro American-246	1,43		
Euro American-34	1,43		
Indo-Oceanic-56	1,43		

Haarlem-42/43/45

Cluster : 42

Euro American-42	93,33
East-African Indian-68	2,22
Euro American-224	2,22
West African II	2,22

Cluster : 43

Euro American-43	96,43
Euro American-15	3,57

Cluster : 45

Euro American-45	86,36
Euro American-190	4,55
Euro American-121	2,27
Euro American-40	2,27
Euro American-42	2,27
Indo-Oceanic-64	2,27
